

POPULATION DENSITY ESTIMATION USING
LINE TRANSECT SAMPLING

BY

JOHN A. ONDRASIK

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1979

To Toni

For Her Love and Support

ACKNOWLEDGMENTS

I would like to thank my adviser, Dr. P. V. Rao, for his guidance and assistance throughout the course of this research. His patience and thoughtful advice during the writing of this dissertation is sincerely appreciated. I would also like to thank Dr. Dennis D. Wackerly for the help and encouragement that he provided during my years at the University of Florida.

Special thanks go to my family for the moral support they provided during the pursuit of this degree. I am especially grateful to my wife, Toni, whose love and understanding made it possible for me to finish this project. Her patience and sacrifices will never be forgotten.

Finally, I want to express my thanks to Mrs. Edna Larrick for her excellent job of typing this manuscript despite the time constraints involved.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	iii
LIST OF TABLES	vi
ABSTRACT	vii
CHAPTER	
I INTRODUCTION	1
1.1 Literature Review	1
1.2 Density Estimation Using Line Transects . .	4
1.3 Summary of Results	9
II DENSITY ESTIMATION USING THE INVERSE SAMPLING PROCEDURE	13
2.1 Introduction	13
2.2 A General Model Based on Right Angle Distances and Transect Length	14
2.2.1 Assumptions	15
2.2.2 Derivation of the Likelihood Function	16
2.3 A Parametric Density Estimate	28
2.3.1 Maximum Likelihood Estimate for D .	28
2.3.2 Unbiased Estimate for D	29
2.3.3 Variance of $\hat{D}_u$	31
2.3.4 Sample Size Determination Using $\hat{D}_u$.	32
2.4 Nonparametric Density Estimate	34
2.4.1 The Nonparametric Model for Estimating D	36
2.4.2 An Estimate for $f_Y(0)$	37
2.4.3 Approximations for the Mean and Variance of $\hat{f}_Y(0)$	40
2.4.4 A Monte Carlo Study	42
2.4.5 The Expected Value and Variance for a Nonparametric Estimate of D.	46
2.4.6 Sample Size Determination Using $\hat{D}_N$.	47

TABLE OF CONTENTS (Continued)

CHAPTER		Page
III	DENSITY ESTIMATION BASED ON A COMBINATION OF INVERSE AND DIRECT SAMPLING	49
	3.1 Introduction	49
	3.2 Gates Estimate	50
	3.2.1 The Mean and Variance of $\hat{D}_g$	54
	3.3 Expected Value of $\hat{D}_{CP}$	57
	3.4 Variance of $\hat{D}_{CP}$	65
	3.5 Maximum Likelihood Justification for $\hat{D}_{CP}$	69
IV	DENSITY ESTIMATION FOR CLUSTERED POPULATIONS	71
	4.1 Introduction	71
	4.2 Assumptions	73
	4.3 General Form of the Likelihood Function	76
	4.4 Estimation of D when $p(\cdot)$ and $h(\cdot)$ Have Specific Forms	79
	4.5 A Worked Example	86
	BIBLIOGRAPHY	90
	BIOGRAPHICAL SKETCH	92

LIST OF TABLES

TABLE		Page
1	Number of animals, N_o , that must be sighted to guarantee the estimate, $\hat{D}_u$, has coefficient of variation, $CV(\hat{D}_u)$	34
2	Forms proposed for the function, $g(y)$	36
3	Results of Monte Carlo Study using $g_1(y) = e^{-10y}$.	45
4	Results of Monte Carlo Study using $g_2(y) = 1-y$. .	45
5	Results of Monte Carlo Study using $g_3(y) = 1-y^2$.	46
6	Number of animals, N_o , that must be sighted to guarantee the estimate $\hat{D}_N$ has coefficient of variation, $CV(\hat{D}_N)$	48

Abstract of Dissertation Presented to the Graduate Council
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

POPULATION DENSITY ESTIMATION USING
LINE TRANSECT SAMPLING

By

John A. Ondrasik

December 1979

Chairman: Pejaver V. Rao
Major Department: Statistics

The use of line transect methods in estimating animal and plant population densities has recently been receiving increased attention in the literature. Many of the density estimates which are currently available are based only on the right angle distances from the sighted objects to a randomly placed transect of known length. This type of sampling, wherein an observer is required to travel along a line transect of some predetermined length, will be referred to as the direct sampling method. In contrast, one can use an inverse sampling plan which will allow the observer to terminate sampling as soon as he sights a prespecified number of animals.

An obvious advantage of an inverse sampling plan is that sampling is terminated as soon as the required number of objects are sighted. A disadvantage is the possibility that sampling may not terminate in any reasonable period of time. Consequently, a third sampling plan, in which sampling stops

as soon as either a prespecified number of objects are sighted or a prespecified length of the transect is traversed, is of practical interest. Such a sampling plan will be referred to as the combined sampling method. The objective of this dissertation is to develop density estimation techniques suitable for both inverse and combined sampling plans.

In Chapter II, both a parametric and a nonparametric estimate for the population density are developed using the inverse sampling approach. We will show that a primary advantage of estimation using inverse sampling is the fact that these estimates can be expressed as the product of two independent random variables. This representation not only enables us to obtain the expected value and variance of our estimates easily, but also leads to a simple criterion for sample size determination.

In Chapter III, we derive a parametric density estimate that is suitable for the combined sampling method. This estimate will be shown to be asymptotically unbiased. An approximation to the variance of this estimate is also provided.

The density estimates developed in Chapters II and III are based on the assumption that the sightings of animals are independent events. In Chapter IV we relax this assumption and develop an estimation procedure using inverse sampling that can be applied to clustered populations--those populations composed of small groups or "clusters" of objects.

CHAPTER I

INTRODUCTION

1.1 Literature Review

Our objective in this dissertation is to examine the problem of density estimation in animal and plant populations. The demand for new and more efficient population density estimates has grown quite rapidly in the past few years. Anderson et al. (1976, p. 1) give a good assessment of the present situation and provide some reasons for the renewed interest in this subject in the following paragraph:

The need to accurately and precisely estimate the size or density of biological populations has increased dramatically in recent years. This has been due largely to ecological problems created by the effects of man's rapidly increasing population. Within the past decade, we have witnessed numerous data gathering activities related to the Environmental Impact Statement (NEPA) process or Biological Monitoring programs. Environmental programs related to phosphate, uranium and coal mining and the extraction of shale oil typically require estimates of the size or density of biological populations. The Endangered Species Act has focused attention on the lack of techniques to estimate population size. It now appears that hundreds of species of plants may be protected under the Act, and, therefore, we will need information on the size of some plant populations. Estimation of the size of biological populations was a major objective of the International Biological Program (IBP) (Smith et al. 1975). Finally, we mention that the ability to estimate population size or density is fundamental to efficient wildlife and habitat management and many important studies in basic ecological research.

The estimation of population size has always been a very interesting and complex problem . For a recent review of the general subject area see Seber (1973). Although many of the methods described in Seber's book are quite useful, they are frequently very expensive and time consuming. Estimation methods based on capture-recapture studies would fall into this category. A further problem with many estimation methods is that they are based on models requiring very restrictive assumptions which severely limit their use in analyzing and interpreting the data. For these reasons and others, line transect sampling schemes are becoming more and more popular. This method of sampling requires an observer to travel along a line transect that has been randomly placed through the area containing the population under study and to record certain measurements whenever a member of the population is sighted. There are several density estimation techniques available using line transect data; however, the full potential is yet to be realized.

Density estimation through line transects is typically practical, rapid and inexpensive for a wide variety of populations. Published references to line transect studies date back to the method used by King (See Leopold, 1933) in the estimation of ruffed grouse populations. Since that time, numerous papers investigating line transect models have appeared, e.g., Webb (1942), Hayne (1949), Robinette et al. (1954), Gates et al. (1968), Anderson and Pospahala (1970),

Sen et al. (1974), Burnham and Anderson (1976) and Crain et al. (1978). Since it is commonly assumed by these authors that the objects being sampled are fixed with respect to the transect, line transect models are best suited for either immotile populations, flushing populations (populations where the animal being observed makes a conspicuous response upon the approach of the observer) or slow moving populations.

Examples of such populations are:

- (i) immotile - birds' nests, dead deer and plants,
- (ii) flushing - grouse, pheasants and quail, and
- (iii) slow moving - desert tortoise and gila monster.

The degree to which line transect methods can be applied to more motile populations, such as deer and hare, will depend on the degree to which the basic assumptions are met. In any case, one should proceed cautiously when using these models for motile populations.

Despite the wide applicability of line transect methods, the estimation problem has only recently begun to receive rigorous treatment and attention from a statistical standpoint. Gates et al. (1968) were the first to develop a density estimation procedure within a standard statistical framework. After making certain assumptions with regard to the probability of sighting an animal located at a given right angle distance from the transect, they rigorously derived a population density estimate. In addition, they were the first authors to provide an explicit form for the approximate sampling variance of their density estimate.

While the assumptions of Gates et al. (1968) concerning the probability of sighting an animal did work well for the ruffed grouse populations they were studying, it is clear that the validity of their assumptions will be quite crucial in establishing the validity their density estimates. If the collected data fail to substantiate their assumptions, large biases could occur in the estimates as seen in Robinette et al. (1974). As a result, Sen et al. (1974) and Pollock (1978) relaxed the assumptions of Gates et al. (1968) by using more general forms for the sighting probability, while Burnham and Anderson (1976) developed a nonparametric approach as a means of providing a more robust estimation procedure.

In the following sections, we will outline the general problem of density estimation using line transects, give our approach to the solution of this problem and summarize the results found in the remainder of this work.

1.2 Density Estimation Using Line Transects

The line transect method is simply a means of sampling from some unknown population of objects that are spatially distributed. In the context of animal or plant population density estimation, these objects take the form of mammals, birds, plants, nests, etc., which are distributed over a particular area of interest. From this point on, our references will always be to animal populations with the understanding that the estimation methods we describe are applicable to all populations which satisfy the necessary assumptions.

In the line transect sampling procedure, a line is randomly placed across an area, A , that contains the unknown population of interest. An observer follows the transect and records one or more of the following three pieces of information for each animal sighted:

- (i) The radial distance, r , from the observer to the animal.
- (ii) The right angle distance, y , from the animal to the line transect.
- (iii) The sighting angle, β , between the line transect and the line joining the observer to the point at which the animal is sighted.

These measurements are illustrated in Figure 1.

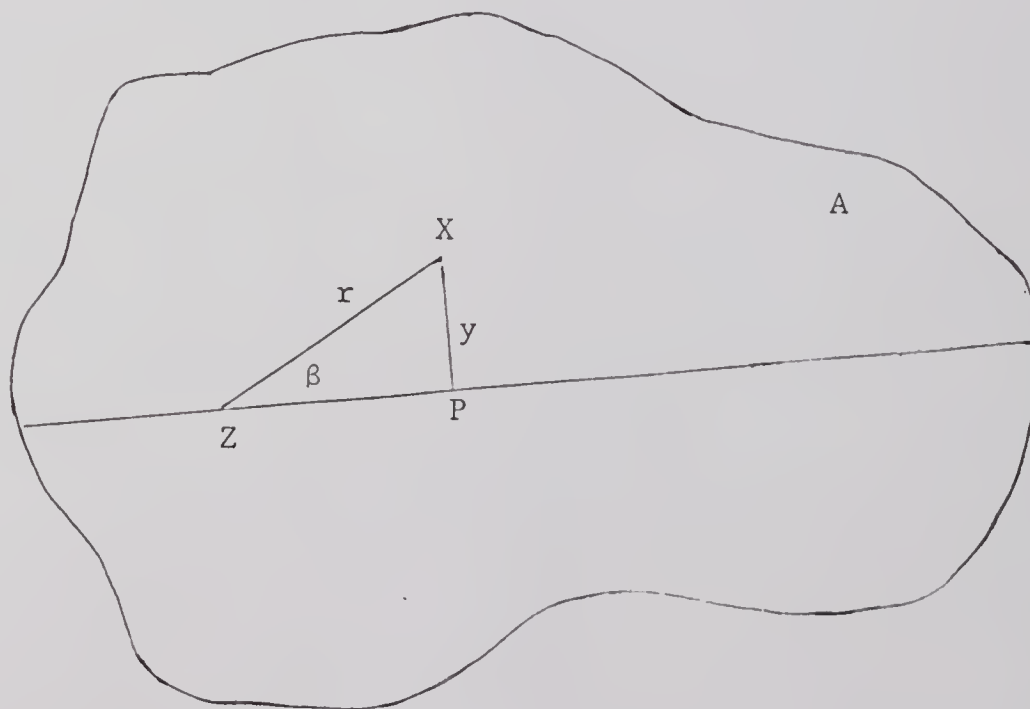


Figure 1. Measurements recorded using line transect sampling. (Z is the position of an observer when an animal is sighted at X . XP is the line from the animal perpendicular to the transect.)

In this work, we shall consider the problem of estimating population density using only the right angle distances. Because estimates depending only on right angle distances are easy and economical to use, such estimates have become very popular over the past several years.

Before any estimation procedure based on right angle distances can be formulated, certain assumptions regarding the population of interest must be made. A set of assumptions used by several workers in the area is detailed in Section 2.2.1. One of the key assumptions in this set is that the probability of sighting an animal located at a right angle distance, y , from the transect can be represented by some nonincreasing function $g(y)$, which satisfies the equality, $g(0) = 1$. This function is simply a mathematical tool for dealing with the fact that animals located closer to the line transect will be seen more readily than animals located further away from the transect. An alternative method of dealing with this phenomenon is given by Anderson and Pospahala (1970).

If $g(y)$ is assumed to have some specific functional form determined by some unknown parameters, then the estimate is said to be parametric. On the other hand, if $g(y)$ is left unspecified except for the requirements that it is nondecreasing and $g(0) = 1$, then the estimate is said to be nonparametric.

Seber (1973) has shown that any density estimate based on right angle distances will have the form

$$\hat{D}_s = \frac{N}{2L_0\hat{c}},$$

where N is a random variable representing the number of animals seen in a line transect of length L_0 and $\hat{c}$ is an estimate for c , a parameter which depends on $g(y)$ through the relation

$$c = \int_0^{\infty} g(y) dy.$$

By noting that the density is simply the number of animals present per unit of area, it is clear that $\hat{c}$ can be interpreted as one-half of the effective width of the strip actually covered by the observer as he moves along the transect. Further examination of $\hat{D}_s$ also points out that estimating the parameter c is the key to the estimation problem.

At this time, we would like to point out that the range for the right angle distance, y , is allowed to go from 0 to $+\infty$, as seen in the integral on the right hand side of the equation for c . In practice, since we are considering only a finite area, A , there will most certainly be a maximum observation distance, W , perpendicular to the transect. However, if W is large enough so that the approximation

$$\int_0^{\infty} g(y) dy \doteq \int_0^W g(y) dy \quad (1.1)$$

is reasonable, then letting y range in the interval $[0, +\infty)$ will not cause any real problems. In practical terms, this means that the probability of observing an animal located beyond the boundary, W , should be essentially zero.

In most real life situations, W can be chosen large enough so that the approximation given in (1.1) is valid. Thus, in the chapters which follow, we will implicitly assume that relation (1.1) holds for the density estimates that we develop.

Both parametric and nonparametric models have been used to derive an estimate for the parameter c , and, consequently, for the population density. In both cases, the estimate for c turns out to be a function of the observed right angle distances. In the parametric case, c will simply be a function of the parameters that define the function chosen for $g(y)$. Examples of parametric estimates are found in Gates et al. (1968), Sen et al. (1974) and Pollock (1978).

Estimation using the nonparametric model is more complicated. Burnham and Anderson (1976) have shown that estimating $\frac{1}{c}$ is equivalent to estimating $f_Y(0)$, where $f_Y(\cdot)$ is the conditional probability density function for right angle distance given an animal is sighted. Thus, the problem of finding a nonparametric estimate for the population density reduces to the problem of estimating a density function at a given point. Unfortunately, this problem has not received much attention in the literature. Burnham and Anderson (1976) suggest four possible estimates for $f_Y(0)$, but the sampling variances associated with these estimates have not been established.

Crain et al. (1978) have also considered the problem of estimating $f_Y(0)$. They derive an estimate using a Fourier Series expansion to approximate the conditional probability density function $f_Y(y)$. Although their procedure does not lead to a simple estimate, they do provide an approximation to its sampling variance.

The line transect method and the corresponding population density estimates so far described require the observer to travel a predetermined distance, L_0 , along the transect. This method will be called the direct sampling method. An alternative to the direct method is the inverse sampling method, wherein sampling is terminated as soon as a specified number, N_0 , of animals are sighted. Clearly, in the direct method, the number of animals seen is a random variable and the total length travelled is a fixed quantity, while in inverse sampling method, the total length travelled is the random variable and the number of animals that must be seen is fixed. The main focus of this work will be to develop density estimation techniques that are based on the inverse sampling method. In addition, we will consider the density estimation problem when a combination of the inverse and direct sampling plans is used.

1.3 Summary of Results

In Chapter 2 we derive two estimates for the population density, D , using an inverse sampling scheme. The set of assumptions which justify the use of these estimates is

similar to those used by Gates et al. (1968) and several others. The estimates have the form

$$\hat{D}_I = \frac{N_o - 1}{2L\hat{c}}$$

where N_o is the number of animals that must be seen before sampling terminates, L is a random variable representing the length travelled on the transect and c is as previously defined. Note the similarity of $\hat{D}_I$ to $\hat{D}_s$ given in Section 1.2. The only difference between the two estimates is that in $\hat{D}_I$ the random variables are L and $\hat{c}$, while in $\hat{D}_s$ they are N and $\hat{c}$. However, this difference gives the inverse sampling method a theoretical advantage over the direct sampling method. The random variables L and $\hat{c}$ will be seen to be independent while N and $\hat{c}$ are not. Thus, the estimate $\hat{D}_I$ is the product of two independent random variables, a fact which not only allows us to obtain its expected value and variance easily, but also leads to a simple criterion for sample size determination.

Both a parametric and a nonparametric estimate for the animal population density are developed in Chapter II. In deriving the parametric estimate, the functional form assumed for $g(y)$ is identical to the one used by Gates et al. (1968). Our parametric density estimate is shown to be unbiased and the exact variance of this estimate is also provided.

In the nonparametric case we propose an estimate for $f_Y(0)$ using the method developed by Loftsgaarden and Quesenberry (1965). We then use heuristic reasons to show

that the corresponding density estimate is asymptotically unbiased, and derive a large sample approximation for its variance.

The inverse sampling method does have one drawback when there is little information available concerning the population to be studied, namely, there exists the possibility that an observer might have to cover a very long transect to sight N_0 animals. To overcome this problem, we develop a parametric density estimate in Chapter III that is based on a combination of the inverse and the direct sampling procedures. In the combined sampling scheme, sampling is terminated when either a prespecified number, N_0 , of animals are sighted or when a prespecified length, L_0 , has been travelled along the transect. Thus, in combined sampling both the length travelled and the number of animals seen will be random variables.

In deriving the density estimate based on the combined sampling method, we again use the functional form for $g(y)$ proposed by Gates et al. (1968). This estimate is shown to be asymptotically unbiased. In addition, an approximate variance for this density estimate is provided.

The density estimates developed in Chapters II and III are based on the assumption that the sightings of animals will be independent events. Gates et al. (1968) showed that this assumption failed to hold for the animal population they were studying. In Chapter IV we relax this assumption, and develop an estimate based on inverse sampling that can be

applied to clustered populations--populations in which the animals aggregate into small groups or "clusters." Since the estimation procedure developed will require the use of a high-speed computer, the last section of Chapter IV is devoted to a worked example to illustrate the computations that would be involved.

CHAPTER II

DENSITY ESTIMATION USING THE INVERSE SAMPLING PROCEDURE

2.1 Introduction

In this chapter we shall propose estimates for animal population density based on an inverse sampling procedure. Unlike the direct sampling method considered by Gates et al. (1968), the inverse sampling procedure specifies the number of animals that must be sighted before the sampling can be terminated. Thus, in the inverse case the number of animals sighted will be a fixed rather than a random quantity. A precise formulation of the inverse sampling method is as follows:

1. Place a line at random across the area, A , to be sampled.
2. Specify a fixed number, N_0 , and sample along the line transect until N_0 animals are observed.

As one proceeds along the transect, certain measurements will be made. These will be denoted by $y_1, y_2, \dots, y_{N_0}$ and ℓ , where y_i is the right angle distance from the i^{th} animal observed to the transect and ℓ is the total distance travelled along the transect during the observation period. A visual depiction of these measurements is given in Figure 2.

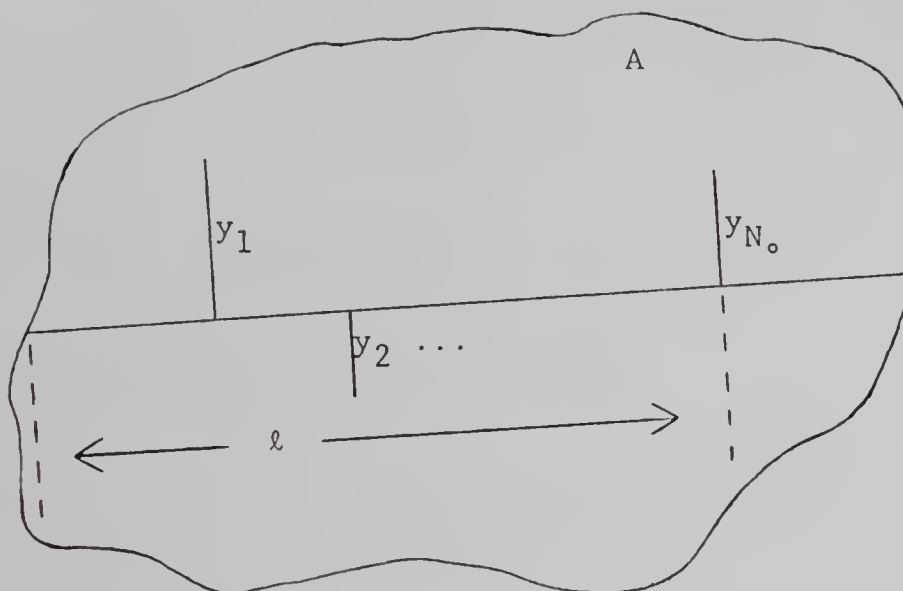


Figure 2. Measurements recorded using inverse sampling.

2.2 A General Model Based on Right Angle Distances and Transect Length

The estimates for the density, D , that we will develop are based on the right angle distances, $y_1, y_2, \dots, y_{N_0}$, and the total distance, ℓ , travelled along the transect. Two possible approaches to the estimation of D merit consideration. First, recall that density is defined as the number of animals present per unit of area, or equivalently the rate at which animals are distributed over some specific area. Therefore, we can write

$$D = \frac{N^*}{A},$$

where A is the area of interest and N^* is the total number of animals present in A . In the direct sampling approach

the estimation of D is most often accomplished by first estimating N^* and then dividing by A . Seber (1973) shows that any estimate of N^* based on direct sampling has the form

$$\hat{N}^* = \frac{NA}{2L_o\hat{c}}$$

where N is a random variable denoting the number of animals seen, L_o is the length of the transect and $\hat{c}$ is an estimate of c , a parameter which depends on the probability of sighting an animal given its right angle distance from the transect. Note that in Seber's estimate, N is random and L_o is fixed. It follows then, that Seber's estimate for D does not depend explicitly on A and has the form

$$\hat{D}_s = \frac{N}{2L_o\hat{c}} .$$

Therefore, the estimate of D is independent of the actual size of A , a property that any reasonable estimate of D should possess.

As an alternative, D itself can be regarded as the basic parameter of interest and estimates for D can be derived directly. This is the approach taken by Burnham and Anderson (1976) and the one that we will follow in developing our estimates.

2.2.1 Assumptions

The form of any estimate of D , the animal population density, will depend upon the type of assumptions we can

make regarding the distribution of the animals to be censused and the nature of the observations that will be made. The assumptions our estimates will be based on are as follows:

- A1. The animals are randomly distributed with rate or density D over the area of interest A , i.e., the probability of a given animal being in a particular region of Area, δA , is $\delta A/A$.
- A2. The animals are independently distributed over A , i.e., given two disjoint regions of area, δA_1 and δA_2 ,

$$P(n_1 \text{ animals are in } \delta A_1 \text{ and } n_2 \text{ animals are in } \delta A_2) \\ = P(n_1 \text{ animals are in } \delta A_1)P(n_2 \text{ animals are in } \delta A_2).$$
- A3. The probability of sighting an animal depends only on its distance from the transect. In addition, there exists a function $g(y)$ giving the conditional probability of observing an animal given its right angle distance, y , from the transect. In probability notation,

$$g(y) = P(\text{observing an animal} \mid y).$$
- A4. $g(0) = 1$, i.e., animals on the line are seen with probability one.
- A5. Animals are fixed, i.e., there is no confusion over animals moving during sampling and none are counted twice.

2.2.2 Derivation of the Likelihood Function

We will use the maximum likelihood procedure to obtain an estimate for D . The joint density function we are interested in is

$$f_{\underline{Y}, L}(\underline{Y}, \ell; N_0)$$

where $\underline{Y} = (Y_1, Y_2, \dots, Y_{N_0})$ is the vector of random variables

representing the right angle distances, L is the random variable representing the total length travelled, and N_0 is the specified number of animals to be seen before sampling terminates. Since the dependence of the joint density on N_0 is implicit throughout the rest of this chapter, it will be dropped from our notation for convenience. Thus, from now on we will denote the density as

$$f_{Y,L}(y,\ell),$$

and all other expressions depending on N_0 in this manner will be handled accordingly.

The following two theorems will be very useful in the derivation of the likelihood function.

Theorem 1: Let $N(\ell)$ denote the number of animals sighted in the interval $(0,\ell]$ along the transect. Then, $N(\ell)$ is a Poisson process, and for some $\theta > 0$,

$$P\{N(\ell)=n\} = \frac{e^{-\theta\ell}(\theta\ell)^n}{n!}, \quad n=0,1,2,\dots$$

Note that the quantity $\theta\ell$ equals the expected number of animals sighted per segment of length ℓ .

Proof: In order to show that $N(\ell)$ is a Poisson process, we will show that the assumptions in Section 2.2.1 imply the postulates necessary for a Poisson process given in Lindgren (1968, p. 162).

First, consider two disjoint intervals, ℓ_1 and ℓ_2 , along the transect and the corresponding areas, $A(\ell_1)$ and $A(\ell_2)$, enclosed by lines perpendicular to the transect as shown in Figure 3.

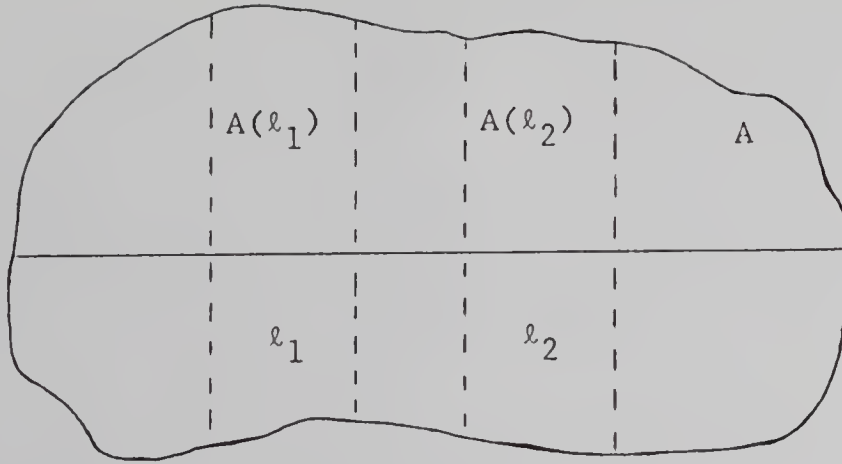


Figure 3. Two disjoint areas along the transect.

Now let N_1 and N_2 be random variables representing the total number of animals that occupy $A(l_1)$ and $A(l_2)$, respectively. By definition, $N(l_1)$ and $N(l_2)$ are the number of animals sighted in $A(l_1)$ and $A(l_2)$, respectively. We know from assumption A2 that N_1 and N_2 are independent, and from assumption A3 that sighting an animal depends only on its distance from the transect. Thus, $N(l_1)$, which depends solely on N_1 and the distances to the N_1 animals from the transect, is independent of $N(l_2)$, i.e., the number of sightings that occur in two disjoint intervals along the transect are independent events.

Next we will show that for every $l > m \geq 0$ and any $h > 0$, $N(l) - N(m)$ and $N(l+h) - N(m+h)$ are identically distributed. First, note that the effective area sampled in seeing $N(l) - N(m)$ animals and $N(l+h) - N(m+h)$ animals is equal to $A(l-m)$ as seen in Figure 4.

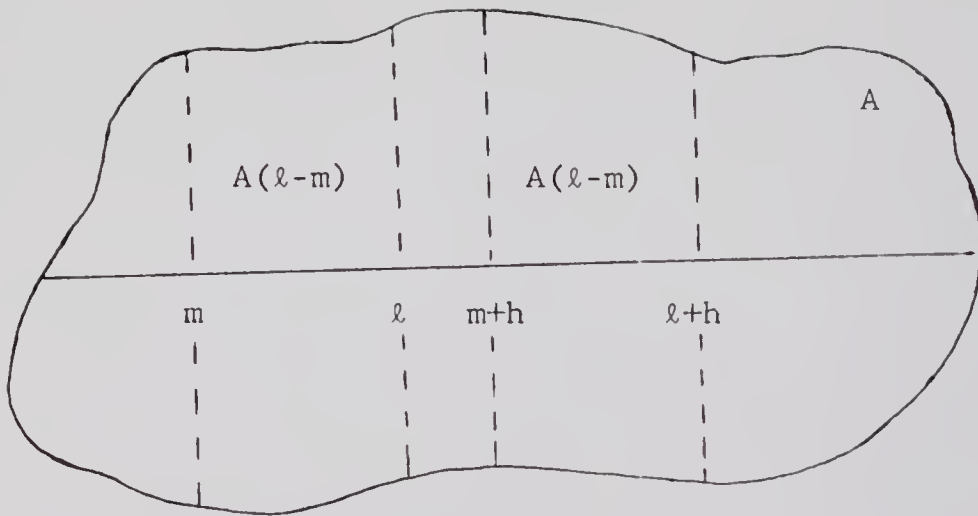


Figure 4. Effective area sampled in seeing $N(l)-N(m)$ animals and $N(l+h)-N(m+h)$ animals.

Therefore, by assumptions A1, A2, and A3, and since the transect is dropped at random, it follows that

$$P\{N(l)-N(m) = j\} = P\{N(l+h)-N(m+h) = j\}, \quad j = 0, 1, 2, \dots$$

Next we must show that for every $l > 0$, and some $\theta > 0$,

$$P\{N(l) = 1\} = \theta l + o(l), \quad \text{as } l \rightarrow 0,$$

where $o(l)$ is a function such that

$$\lim_{l \rightarrow 0} \frac{o(l)}{l} = 0.$$

Again let $A(l)$ be the area defined by l on the transect. Now define B_j to be the event $\{N(l) = j\}$ and E_j to be the event that there are exactly j animals in area $A(l)$. Then it follows that

$$\begin{aligned} P(B_1) &= \sum_{j=1}^{\infty} P(B_1 E_j) \\ &= \sum_{j=1}^{\infty} P(B_1 | E_j) P(E_j). \end{aligned}$$

Under assumptions A1 and A2, Pielou (1969, p. 81) has shown that

$$P(E_j) = \frac{e^{-DA(\ell)} [DA(\ell)]^j}{j!}, \quad j = 0, 1, 2, \dots$$

Also, under assumptions A1, A2 and A3, Seber (1973, Eq. (2.6)) has shown that

$$P(B_1 | E_1) = \frac{2\ell c}{A(\ell)},$$

where

$$c = \int_0^\infty g(y) dy. \quad (2.1)$$

Therefore, we can write

$$P(B_1) = 2cD\ell e^{-DA(\ell)} + \sum_{j=2}^{\infty} P(B_1 | E_j) P(E_j),$$

and if we show

$$\sum_{j=2}^{\infty} P(B_1 | E_j) P(E_j) = o(\ell)$$

the proof will be complete. Note that

$$\begin{aligned} \sum_{j=2}^{\infty} P(B_1 | E_j) P(E_j) &\leq \sum_{j=2}^{\infty} \frac{e^{-DA(\ell)} [DA(\ell)]^j}{j!} \\ &= DA(\ell) \sum_{j=2}^{\infty} \frac{e^{-DA(\ell)} [DA(\ell)]^{j-1}}{j!} \\ &\leq DA(\ell) \sum_{j=2}^{\infty} \frac{e^{-DA(\ell)} [DA(\ell)]^{j-1}}{(j-1)!} \\ &= DA(\ell) [1 - e^{-DA(\ell)}]. \end{aligned}$$

For any finite area A , $A(\ell)$ is $O(\ell)$, that is

$$\lim_{\ell \rightarrow 0} \frac{A(\ell)}{\ell} < K, \text{ for some } K > 0.$$

Therefore, as $\ell \rightarrow 0$

$$\sum_{j=2}^{\infty} P(B_1 | E_j) P(E_j) = o(\ell).$$

and, upon writing

$$\theta = 2cD, \quad (2.2)$$

we get, as $\ell \rightarrow 0$,

$$P(B_1) = \theta\ell + o(\ell).$$

Finally, we need to show that for every $\ell > 0$,

$$\sum_{n>1} P\{N(\ell) = n\} = o(\ell), \text{ as } \ell \rightarrow 0.$$

Note that for all $n > 1$, we can write

$$\begin{aligned} P(B_n) &= \sum_{j=1}^{\infty} P(B_n | E_j) \\ &= \sum_{j=n}^{\infty} P(B_n | E_j) P(E_j). \end{aligned}$$

Again, by using the fact that $A(\ell)$ is $O(\ell)$, it is easy to show that

$$P(B_n) = o(\ell), \text{ as } \ell \rightarrow 0,$$

and $N(\ell)$ satisfies the four conditions necessary for a Poisson process.

Before proceeding to the second theorem, we need to define the following random variables. Let T_i denote the random

variable corresponding to the distance travelled on the transect between sightings of the $(i-1)^{\text{st}}$ and i^{th} animals, $i = 1, 2, \dots, N_0$. Then the total distance travelled is given by

$$L = \sum_{i=1}^{N_0} T_i.$$

The following theorem establishes the independence of $\underline{Y}$ and $T_1, T_2, \dots, T_{N_0}$, for the case $N_0 = 2$, and this fact enables us to derive the joint density function, $f_{\underline{Y}, L}(\underline{y}, \ell)$.

Theorem 2. The random variables T_1 , T_2 , Y_1 and Y_2 are mutually independent.

Proof: In order to establish the independence of T_1 , T_2 , Y_1 and Y_2 we will derive the joint density

$$f_{T_1, T_2, Y_1, Y_2}(t_1, t_2, y_1, y_2)$$

and show that it can be factored into four functions, each depending on only one of the random variables of interest.

Let y_1 , y_2 , t_1 , t_2 , h_1 , h_2 , g_1 and g_2 be non-negative real numbers such that

$$t_1 + h_1 < t_1 + t_2,$$

$$y_1 + g_1 < y_2$$

as shown in Figure 5.

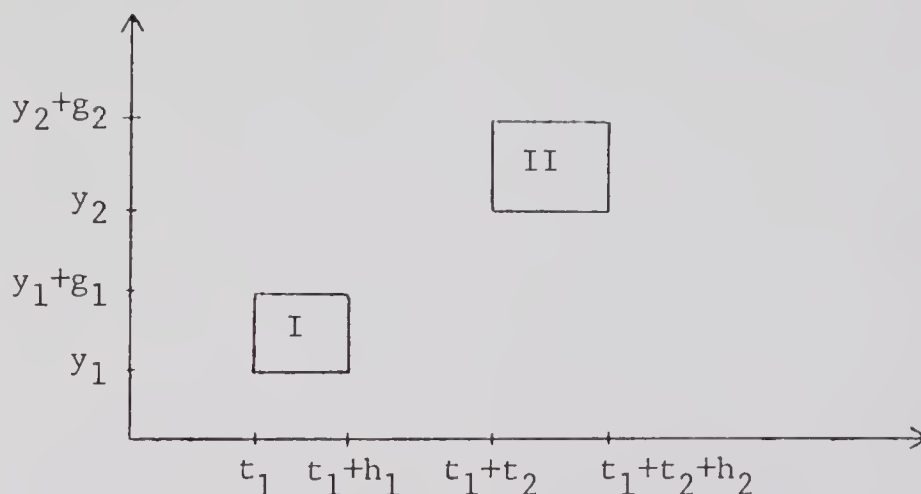


Figure 5. Areas defined by $y_1, y_2, t_1, t_2, g_1, g_2, h_1$, and h_2 .

Now let

$$P(h_1, g_1, h_2, g_2) = P(t_1 < T_2 \leq t_1 + h_1, y_1 < Y_1 \leq y_1 + g_1, \\ t_2 < T_2 \leq t_2 + h_2, y_2 < Y_2 \leq y_2 + g_2).$$

Then we can write

$$f_{T_1, T_2, Y_1, Y_2}(t_1, t_2, y_1, y_2) = \lim_{\substack{h_i \rightarrow 0 \\ g_i \rightarrow 0 \\ i = 1, 2}} \frac{P(h_1, g_1, h_2, g_2)}{h_1 g_1 h_2 g_2},$$

provided the limit exists.

Now notice that the event whose probability we wish to find, namely

$$\{t_1 < T_1 \leq t_1 + h_1, y_1 < Y_1 \leq y_1 + g_1, t_2 < T_2 \leq t_2 + h_2, y_2 < Y_2 \leq y_2 + g_2\},$$

is equivalent to the intersection of the following events:

S_1 , the event $\{N(t_1) = 0\}$

S_2 , the event $\{N(t_1 + h_1) - N(t_1) = 1\}$ and $\{y_1 < Y_1 \leq y_1 + g_1\}$,

i.e., an animal is seen in area I

S_3 , the event $\{N(t_1+t_2) - N(t_1+h_1) = 0\}$

S_4 , the event $\{N(t_1+t_2+h_2) - N(t_1+t_2) = 1\}$ and

$\{y_2 < Y_2 < y_2 + g_2\}$, i.e., an animal is seen in area II.

Now, by Theorem 1 and Assumption A3, the events S_1 , S_2 , S_3 and S_4 are independent so that we can write

$$\begin{aligned} P(h_1, g_1, h_2, g_2) &= P(S_1 S_2 S_3 S_4) \\ &= P(S_1) P(S_2) P(S_3) P(S_4). \end{aligned}$$

We now need to find expressions for the probabilities of S_1 , S_2 , S_3 and S_4 . Since $N(t)$ is a Poisson process,

$$P(S_1) = e^{-\theta t_1},$$

and

$$P(S_3) = e^{-\theta(t_2 - h_1)}.$$

However, $P(S_2)$ and $P(S_4)$ are not so easily obtained. We will only show how to find $P(S_2)$, since $P(S_4)$ is found in a similar fashion. First, define S_{2j} to be the event that there are exactly j animals in area I. Then

$$\begin{aligned} P(S_2) &= \sum_{j=1}^{\infty} P(S_2 S_{2j}) \\ &= \sum_{j=1}^{\infty} P(S_2 | S_{2j}) P(S_{2j}). \end{aligned}$$

By assumptions A1 and A2, the number of animals located in area I will be distributed as a Poisson random variable with parameter $2Dg_1h_1$, where D is the density of the animals (see Pielou, 1969, p. 81). Note, the factor of 2 comes in since area I can be found on both sides of the transect. Therefore,

$$P(S_{2j}) = \frac{e^{-2Dg_1h_1} (2Dg_1h_1)^j}{j!}, \quad j = 0, 1, 2, \dots$$

By assumption A3,

$$P(S_2|S_{21}) = g(y'_1),$$

for some $y_1 < y'_1 \leq y_1 + g_1$. Therefore, by an argument similar to that found in Theorem 1, as $g_1 \rightarrow 0$ and $h_1 \rightarrow 0$,

$$\begin{aligned} P(S_2) &= 2Dg_1h_1 e^{-2Dg_1h_1} g(y'_1) + \sum_{j=2}^{\infty} P(S_2|S_{2j})P(S_{2j}) \\ &= 2Dg_1h_1 e^{-2Dg_1h_1} g(y'_1) + o(g_1h_1). \end{aligned}$$

Similarly, we can show that as $g_2 \rightarrow 0$ and $h_2 \rightarrow 0$,

$$P(S_4) = 2Dg_2h_2 e^{-2Dg_2h_2} g(y'_2) + o(g_2h_2),$$

for some $y_2 < y'_2 \leq y_2 + g_2$. Thus, we can substitute the expressions for $P(S_1)$, $P(S_2)$, $P(S_3)$ and $P(S_4)$ into $P(h_1, g_1, h_2, g_2)$ to obtain

$$\begin{aligned} P(h_1, g_1, h_2, g_2) &= e^{-\theta t_1} e^{-\theta(t_2 - h_1)} \{2Dg_1h_1 e^{-2Dg_1h_1} g(y'_1) + o(g_1h_1)\} \\ &\quad \times \{2Dg_2h_2 e^{-2Dg_2h_2} g(y'_2) + o(g_2h_2)\}. \end{aligned}$$

Consequently,

$$\lim_{\substack{h_i \rightarrow 0 \\ g_i \rightarrow 0 \\ i=1,2}} \frac{P(h_1, g_1, h_2, g_2)}{g_1 h_1 g_2 h_2} = 4D^2 e^{-\theta t_1} e^{-\theta t_2} g(y_1) g(y_2), \quad (2.3)$$

which completes the proof of Theorem 2.

In the same manner, we can show that the independence established in Theorem 2 will hold for any finite number of sightings, N_o . In this case if $\underline{T} = (T_1, T_2, \dots, T_{N_o})$ and $\underline{Y} = (Y_1, Y_2, \dots, Y_{N_o})$, then (2.3) becomes

$$f_{\underline{T}, \underline{Y}}(\underline{t}, \underline{y}) = 2^{N_o D^{N_o}} e^{-\theta \sum_{i=1}^{N_o} t_i} \prod_{i=1}^{N_o} g(y_i).$$

Upon using equation (2.2) in $f_{\underline{T}, \underline{Y}}(\underline{t}, \underline{y})$, we get

$$f_{\underline{T}, \underline{Y}}(\underline{t}, \underline{y}) = \theta^{N_o} e^{-\theta \sum_{i=1}^{N_o} t_i} c^{-N_o} \prod_{i=1}^{N_o} g(y_i).$$

Thus, the marginal distributions for T_i and Y_i are

$$f_{Y_i}(y_i) = \frac{g(y_i)}{c}, \quad y_i > 0,$$

and

$$f_{T_i}(t_i) = \theta e^{-\theta t_i}, \quad t_i > 0.$$

Therefore, $T_1, T_2, \dots, T_{N_o}$ are independent, identically distributed (iid) as Exponential random variables with parameter θ , and

$$L = \sum_{i=1}^{N_o} T_i$$

has a Gamma distribution with parameters N_o and θ , i.e.,

$$f_L(\ell) = \frac{\theta^{N_o} \ell^{N_o-1} e^{-\theta \ell}}{\Gamma(N_o)} \quad \ell > 0, \quad \theta > 0.$$

Furthermore, L is independent of $\underline{Y}$.

The likelihood function for the estimation of θ and c can now be obtained by taking the product of $f_L(\ell)$ and $f_Y(\underline{y})$, i.e.,

$$L(\theta, c; \underline{y}, \ell) = \frac{1}{c^{N_0}} \prod_{i=1}^{N_0} g(y_i) \frac{\theta^{N_0} \ell^{N_0-1} e^{-\theta \ell}}{\Gamma(N_0)} . \quad (2.4)$$

We will now outline how one can estimate D , the animal population density, from the likelihood function given in (2.4). As noted earlier, D is related to θ and c by equation (2.2), i.e.,

$$D = \frac{\theta}{2c} .$$

Thus, the maximum likelihood estimate for D would be

$$\hat{D} = \frac{\hat{\theta}}{2\hat{c}}$$

where $\hat{\theta}$ and $\hat{c}$ are maximum likelihood estimates of θ and c , respectively, obtained from (2.4). Note that the estimate $\hat{D}$ is the ratio of two mutually independent random variables, one depending on L alone and the other depending on $\underline{Y}$ alone. This property will be found to be very useful when evaluating the moments of $\hat{D}$.

We have now set the framework necessary for deriving an estimate of D . In the next section we shall obtain an estimate for D assuming that $g(y)$ has a particular parametric form.

2.3 A Parametric Density Estimate

Any estimate for D that is derived after assuming an explicit function for $g(y)$ will be called a parametric estimate. Gates et al. (1968), using direct sampling, derived an estimate for D assuming

$$g(y) = e^{-\lambda y}.$$

Using this same function for $g(y)$, we will derive the corresponding estimate based on inverse sampling.

2.3.1 Maximum Likelihood Estimate for D

To estimate D we need to estimate both θ and c from the likelihood function (2.4). In this case

$$g(y) = e^{-\lambda y}, \quad y > 0, \lambda > 0$$

so that

$$c = \frac{1}{\lambda}.$$

Substituting for c in (2.2) yields

$$D = \frac{\theta \lambda}{2}. \quad (2.5)$$

Also, by substituting for c in (2.4), the likelihood function becomes

$$L(\theta, \lambda; \underline{y}, \ell) = \lambda^{N_0} e^{-\lambda \sum_{i=1}^{N_0} y_i} \frac{0^{N_0 \ell} N_0 - 1 e^{-0 \ell}}{\Gamma(N_0)}, \quad \ell > 0, y_i > 0. \quad (2.6)$$

The joint maximum likelihood estimates for θ and λ can now be easily obtained. The natural logarithm of the likelihood function is

$$\ln L(\theta, \lambda; \underline{y}, \ell) = N_o \ln \lambda - \lambda \sum_{i=1}^{N_o} y_i + N_o \ln \theta + (N_o - 1) \ln \ell - \theta \ell - \ln \Gamma(N_o).$$

Taking the partial derivatives with respect to θ and λ yields

$$\frac{\partial \ln L(\theta, \lambda; \underline{y}, \ell)}{\partial \theta} = \frac{N_o}{\theta} - \ell,$$

$$\frac{\partial \ln L(\theta, \lambda; \underline{y}, \ell)}{\partial \lambda} = \frac{N_o}{\lambda} - \sum_{i=1}^{N_o} y_i.$$

Setting these equal to 0 yields

$$\hat{\theta} = \frac{N_o}{\ell},$$

and

$$\hat{\lambda} = \frac{N_o}{\sum_{i=1}^{N_o} y_i}.$$

Substituting these estimates for θ and λ in (2.5), the maximum likelihood estimate for D is seen to be

$$\hat{D} = \frac{\hat{\theta} \hat{\lambda}}{2} = \frac{N_o^2}{2 \ell \sum_{i=1}^{N_o} y_i}.$$

2.3.2 Unbiased Estimate for D

The expected value of the estimate D , developed in Section 2.3.1, is

$$\begin{aligned} E(\hat{D}) &= E\left(\frac{\hat{\theta} \hat{\lambda}}{2}\right) \\ &= \frac{1}{2} E(\hat{\theta}) E(\hat{\lambda}) \end{aligned}$$

since $\hat{\theta}$ and $\hat{\lambda}$ are independent. Using the fact that L has a Gamma distribution with parameters N_o and θ , we obtain

$$\begin{aligned} E(\hat{\theta}) &= E\left(\frac{N_o}{L}\right) \\ &= \frac{N_o \theta}{(N_o - 1)} . \end{aligned}$$

To derive an expression for $E(\hat{\lambda})$, first recall that $Y_1, \dots, Y_{N_o}$ are iid with the common density

$$f_Y(y) = \frac{g(y)}{c} = \lambda e^{-\lambda y}, \quad y > 0.$$

Therefore, $\sum_{i=1}^{N_o} Y_i$ is distributed as a Gamma random variable with parameters N_o and λ and

$$\begin{aligned} E(\hat{\lambda}) &= E\left[\frac{N_o}{\sum_{i=1}^{N_o} Y_i}\right] \\ &= \frac{N_o \lambda}{(N_o - 1)} . \end{aligned}$$

Independence of $\hat{\theta}$ and $\hat{\lambda}$ now yields

$$\begin{aligned} E(\hat{D}) &= \frac{1}{2} E(\hat{\theta}) E(\hat{\lambda}) \\ &= \frac{N_o^2}{(N_o - 1)^2} \frac{\lambda \theta}{2} \\ &= \frac{N_o^2}{(N_o - 1)^2} D . \end{aligned}$$

An unbiased estimate for D is, therefore, given by

$$\begin{aligned} \hat{D}_u &= \frac{(N_o - 1)^2}{N_o^2} \frac{\hat{\theta} \hat{\lambda}}{2} \\ &= \frac{(N_o - 1)^2}{2L \sum_{i=1}^{N_o} Y_i} . \end{aligned} \tag{2.7}$$

2.3.3 Variance of $\hat{D}_u$

Due to the independence of L and $\sum_{i=1}^{N_o} Y_i$, the variance of $\hat{D}_u$ can be derived directly. We have

$$\begin{aligned} \text{Var}(\hat{D}_u) &= \text{Var} \left[\frac{(N_o - 1)^2}{N_o \sum_{i=1}^{N_o} Y_i} \right] \\ &= \frac{(N_o - 1)^4}{4} \text{Var} \left[\frac{1}{L \sum_{i=1}^{N_o} Y_i} \right] \\ &= \frac{(N_o - 1)^4}{4} E \left[\frac{1}{L^2 \left(\sum_{i=1}^{N_o} Y_i \right)^2} - \left(E \left[\frac{1}{L \sum_{i=1}^{N_o} Y_i} \right] \right)^2 \right] \end{aligned}$$

Since L and $\sum_{i=1}^{N_o} Y_i$ are independent it follows that

$$\text{Var}(\hat{D}_u) = \frac{(N_o - 1)^4}{4} \left[E \left(\frac{1}{L^2} \right) E \left[\frac{1}{\left(\sum_{i=1}^{N_o} Y_i \right)^2} \right] - E^2 \left(\frac{1}{L} \right) E^2 \left(\frac{1}{\sum_{i=1}^{N_o} Y_i} \right) \right]. \quad (2.8)$$

Deriving the $\text{Var}(\hat{D}_u)$ now reduces to the problem of evaluating

the expected values for $\frac{1}{L}$, $\frac{1}{\sum_{i=1}^{N_o} Y_i}$, $\frac{1}{L^2}$, and $\frac{1}{\left(\sum_{i=1}^{N_o} Y_i \right)^2}$. Expressions for these quantities are easily obtained by noting that

L and $\sum_{i=1}^{N_o} Y_i$ have Gamma distributions with respective parameters

$N_o, 0$ and N_o, λ . Straightforward calculations show that for $N_o > 2$,

$$E\left(\frac{1}{L}\right) = \frac{\theta}{(N_o-1)} \quad (2.9)$$

$$E\left(\frac{1}{\sum_{i=1}^{N_o} Y_i}\right) = \frac{\lambda}{(N_o-1)} \quad (2.10)$$

$$E\left(\frac{1}{L^2}\right) = \frac{\theta^2}{(N_o-1)(N_o-2)} \quad (2.11)$$

$$E\left[\frac{1}{\left(\sum_{i=1}^{N_o} Y_i\right)^2}\right] = \frac{\lambda^2}{(N_o-1)(N_o-2)} \quad (2.12)$$

Now using (2.5), (2.9), (2.10), (2.11), and (2.12) in (2.8) we get

$$\begin{aligned} \text{Var}(\hat{D}_u) &= \frac{(N_o-1)^4}{4} \left\{ \frac{\theta^2 \lambda^2}{(N_o-1)^2 (N_o-2)^2} - \frac{\theta^2 \lambda^2}{(N_o-1)^4} \right\} \\ &= \frac{(N_o-1)^2 \theta^2 \lambda^2}{4} \left\{ \frac{1}{(N_o-2)^2} - \frac{1}{(N_o-1)^2} \right\} \\ &= D^2 \frac{(2N_o-3)}{(N_o-2)^2}, \end{aligned} \quad (2.13)$$

provided $N_o > 2$. Note that $\text{Var}(\hat{D}_u)$ does not exist if $N_o \leq 2$.

2.3.4 Sample Size Determination Using $\hat{D}_u$

The first problem in designing a survey using the inverse line transect method is to determine in advance the number of animals, N_o , that must be sighted before sampling terminates. One criterion for the selection of N_o (see Seber, 1973) is the requirement that the design must yield an estimate of the density, D , with a prescribed coefficient of variation,

$$CV = \frac{\sigma_{\hat{D}}}{E(\hat{D})} ,$$

where $\sigma_{\hat{D}}$ and $E(\hat{D})$ denote, respectively, the standard deviation and the expected value for the estimate, $\hat{D}$. As one can see immediately, small values of CV are desirable since this indicates that the estimate has a small standard deviation relative to its expected value.

With the inverse sampling method, the value of N_0 needed to guarantee a preset value, C, for the coefficient of variation of $\hat{D}_u$ can be calculated easily. Using (2.7) and (2.13) we see that, for $N_0 > 2$

$$CV(\hat{D}_u) = \frac{(2N_0 - 3)^{1/2}}{N_0 - 2} .$$

Then, setting $C = CV(\hat{D}_u)$, it is easily shown that N_0 is the root of the quadratic equation

$$C^2 N_0^2 - (4C^2 + 2)N_0 + 4C^2 + 3 = 0 .$$

Solving for N_0 yields the two roots

$$N_0 = 2 + \frac{1 \pm (1 + C^2)^{1/2}}{C^2} .$$

Since the variance of $\hat{D}_u$ exists only for $N_0 > 2$, the required sample size is

$$N_0 = 2 + \frac{1 + (1 + C^2)^{1/2}}{C^2} .$$

For example, if $C = .25$, then $N_0 = 35$. Table 1 gives values of N_0 corresponding to coefficients of variation ranging from .1 to .5.

Table 1. Number of animals, N_0 , that must be sighted to guarantee the estimate, $\hat{D}_u$, has coefficient of variation, $CV(\hat{D}_u)$.

$CV(\hat{D}_u)$	N_0
.50	11
.40	15
.30	25
.25	35
.20	53
.15	92
.10	203

2.4 Nonparametric Density Estimate

In this section we will consider a nonparametric estimate for the population density, D , using inverse sampling. In contrast to the parametric approach used in Section 2.3, the nonparametric approach leaves the function $g(y)$, which represents the probability of observing an animal given its right angle distance, unspecified.

In Section 2.2.2 we showed that an estimate for D is given by

$$\hat{D} = \frac{\hat{\theta}}{2\hat{c}},$$

where $\hat{\theta}$ and $\hat{c}$ are the estimates for θ , the expected number of sightings per unit length of the transect, and c defined as

$$c = \int_0^{\infty} g(y)dy.$$

If $g(y)$ is completely specified, except for some parameters, then the problem of estimating D reduces to the problem of estimating θ and the parameters in $g(y)$. In Section 2.3 we considered the specific case

$$g(y) = e^{-\lambda y}.$$

A drawback to this approach, where we specify a functional form for $g(y)$, is that the function chosen must take into account the inherent detection difficulties that are present when a particular animal species is being sampled. If one examines the various forms that have been suggested for $g(y)$, one quickly becomes aware of the problem of finding a form that is flexible enough to accommodate the many possibilities which exist. Some of the functions that have been proposed for $g(y)$ are presented in Table 2. As seen in the table, the suggestions for $g(y)$ represent a number of different shapes in an effort to reflect the nature of the animal being sampled and the type of ground cover being searched.

Because of the problems that can arise in choosing a function for $g(y)$, Burnham and Anderson (1976) considered a nonparametric approach as a means of avoiding the need for the specification of $g(y)$. Leaving $g(y)$ unspecified will allow the estimation procedure to depend on the observations that are actually made, not on any particular model. Thus, a nonparametric model might provide a more robust estimation method, that is, an estimation method that could be applied to a much wider class of animal species.

Table 2. Forms proposed for the function, $g(y)$.

Function	Author
$g(y) = e^{-\lambda y}, \lambda > 0$	Gates et al. (1968)
$g(y) = \begin{cases} 1 - \left(\frac{y}{w}\right)^a, & 0 < y < w \\ 0 & y > w \end{cases}$	Eberhardt (1968)
$g(y) = \begin{cases} 1, & 0 < y < w \\ 0, & y > w \end{cases}$	Seber (1973)
$g(y) = \int_y^{\infty} \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} dx, \beta > 0, \alpha > 0$	Sen et al. (1974)
$g(y) = e^{-\left(\frac{y}{\lambda}\right)^p}, p > 0, \lambda > 0$	Pollock (1978)

2.4.1 The Nonparametric Model for Estimating D

Consider the estimate for D developed in Section 2.2.2, that is

$$\hat{D} = \frac{\hat{\theta}}{2\hat{c}}.$$

As noted earlier, if

$$g(y) = e^{-\lambda y}$$

then

$$c = \frac{1}{\lambda},$$

and our estimate for D is

$$\hat{D} = \frac{\hat{\theta}\hat{\lambda}}{2}.$$

Now, if $g(y)$ is left unspecified, then an estimate for $\frac{1}{c}$ may be obtained along the same lines Burnham and Anderson (1976) used in the case of direct sampling. By assumption A4,

$$f_Y(0) = \frac{g(0)}{c} = \frac{1}{c}.$$

Hence, $\frac{1}{c}$ equals the value of the $f_Y(\cdot)$ evaluated at $y=0$, where $f_Y(\cdot)$ is the probability density function for the right angle distance, Y , given an animal is seen. The problem of finding a nonparametric estimate for $\frac{1}{c}$, therefore, reduces to the problem of finding an estimate, $\hat{f}_Y(0)$, for $f_Y(0)$.

An estimate for D will then be given by

$$\hat{D} = \frac{\hat{\theta} \hat{f}_Y(0)}{2}, \quad (2.14)$$

where $\hat{\theta}$ may be taken as the maximum likelihood estimate derived in Section 2.3.1. That is,

$$\hat{\theta} = \frac{(N_o - 1)}{L},$$

where we have replaced N_o by $N_o - 1$ to remove the bias.

2.4.2 An Estimate for $f_Y(0)$

Burnham and Anderson (1976) suggested four possible methods for estimating $f_Y(0)$, but we are not aware of any work which investigates the theoretical properties of any of these estimates. Loftsgaarden and Quesenberry (1965) considered a density function estimate based on the observation that

$$f_Y(x) = \lim_{h \rightarrow 0} \frac{F_Y(x+h) - F_Y(x-h)}{2h}$$

where $F_Y(\cdot)$ is the cumulative distribution function. For the purpose of estimating $f_Y(0)$, their estimate takes the form

$$\hat{f}_Y(0) = \{\sqrt{N_0} Y_{(\lfloor \sqrt{N_0} + 1 \rfloor)}\}^{-1}, \quad (2.15)$$

where $\lfloor \sqrt{N_0} + 1 \rfloor$ is the value of $\sqrt{N_0} + 1$ rounded off to the nearest integer and $Y_{(j)}$ is the j^{th} order statistic of the sample $y_1, y_2, \dots, y_n$.

Loftsgaarden and Quesenberry (1965) showed that $\hat{f}_Y(0)$ as given in (2.15) is a consistent estimate, provided $f_Y(\cdot)$ is a positive and continuous probability density function. One nice property of $\hat{f}_Y(0)$ is that it can be easily calculated from the data. However, evaluation of the moments of this estimate does present some problems. In fact, the mean and the variance may not even exist in some cases. But, whenever $\lfloor \sqrt{N_0} + 1 \rfloor \geq 3$, i.e., whenever $N_0 \geq 4$, the variance of $\hat{f}_Y(0)$ is finite as shown in the following theorem.

Theorem 3. Let $Y_1, Y_2, \dots, Y_n$ be a set of independent, identically distributed random variables, representing the right angle distances, with continuous probability density function (p.d.f.)

$$f_Y(y) = \frac{g(y)}{c}, \quad y > 0.$$

Also, let $Y_{(r)}$ be the r^{th} order statistic. Then

$$E \left[\frac{1}{Y_{(r)}^2} \right] < +\infty$$

for every integer r , such that $3 \leq r \leq n$.

Proof: The density function for $Y_{(r)}$ is

$$h_r(y) = n \binom{n-1}{r-1} F_Y^{r-1}(y) [1-F_Y(y)]^{n-r} f_Y(y)$$

where

$$F_Y(y) = \int_0^y f_Y(t) dt.$$

Therefore,

$$E\left(\frac{1}{Y_{(r)}^2}\right) = n \binom{n-1}{r-1} \int_0^\infty \frac{1}{y^2} F_Y^{r-1}(y) [1-F_Y(y)]^{n-r} dF_Y(y).$$

Since $g(y)$ represents a probability, $g(y) \leq 1$ and

$$F_Y(y) = \int_0^y \frac{g(t)}{c} dt \leq \frac{y}{c}.$$

Therefore,

$$\begin{aligned} E\left(\frac{1}{Y_{(r)}^2}\right) &\leq \frac{n \binom{n-1}{r-1}}{c^2} \int_0^\infty F_Y^{r-3}(y) [1-F_Y(y)]^{n-r} dF_Y(y), \\ &= \frac{n \binom{n-1}{r-1}}{c^2} \frac{\Gamma(r-2) \Gamma(n-r+1)}{\Gamma(n-1)}, \end{aligned}$$

which completes the proof.

Simple asymptotic approximations for the mean and variance of $\hat{f}_Y(0)$ which work well for several densities given in Table 2 can be developed using the first order terms in a formal Taylor series expansion of $\hat{f}_Y(0)$. The basic ideas involved in the derivation of these approximations are presented in the following section.

2.4.3 Approximations for the Mean and Variance of $\hat{f}_Y(0)$

Let $F(\cdot)$ and $F^{-1}(\cdot)$ denote the cumulative distribution function and its inverse for the random variable Y , the right angle distance. Also, let

$$r = [\sqrt{N} + 1],$$

$$U_r = F(Y_{(r)}),$$

and

$$\phi(U_r) = \frac{1}{(r-1)F^{-1}(U_r)}$$

where $Y_{(r)}$ is the r^{th} order statistic in a random sample of size N from $F(\cdot)$. Then proceeding as in Lindgren (1968, p. 409) it is easy to see that

$$\hat{f}_Y(0) = \phi(U_r),$$

$$E(U_r) = \frac{r}{N+1} = p_r,$$

and

$$\text{Var}(U_r) = \frac{p_r(1-p_r)}{N+2}. \quad (2.16)$$

Assuming that $\phi(\cdot)$ is continuous and differentiable once at p_r , the first order terms in the Taylor series expansion of $\phi(\cdot)$ at p_r yields the approximation

$$\phi(U_r) \doteq \phi(p_r) + (U_r - p_r) \left. \frac{d\phi(u)}{du} \right|_{u=p_r} \quad (2.17)$$

Taking expectations on both sides of (2.17) yields

$$E\{\phi(U_r)\} \doteq \phi(p_r),$$

and substituting for r , p_r and $g(\cdot)$ yields

$$E\{\hat{f}_Y(0)\} \doteq \frac{1}{\sqrt{N} F^{-1}\left(\frac{\sqrt{N}+1}{N+1}\right)}.$$

Taking the limit as N tends to infinity and noting that $F^{-1}(0)$ is 0 and $u = F(y)$ yields

$$\begin{aligned} \lim_{N \rightarrow \infty} E\{\hat{f}_Y(0)\} & \doteq \lim_{N \rightarrow \infty} \left[\frac{F^{-1}\left(\frac{\sqrt{N}+1}{N+1}\right)}{\frac{1}{\sqrt{N}}} \right]^{-1} \\ & = \frac{1}{\left. \frac{dF^{-1}(u)}{du} \right|_{u=0}}, \\ & = \left. \frac{dF(y)}{dy} \right|_{y=0} = f_Y(0). \end{aligned} \quad (2.18)$$

Thus for large N , $\hat{f}_Y(0)$ is approximately unbiased.

An approximation for the variance of $\hat{f}_Y(0)$ is found in a similar fashion. Using (2.17) we get

$$\text{Var}\{\phi(U_r)\} \doteq \left\{ \left. \frac{d\phi(u)}{du} \right|_{u=p_r} \right\}^2 \text{Var}(U_r).$$

Evaluating the derivative yields

$$\left. \frac{d\phi(u)}{du} \right|_{u=p_r} = - \frac{1}{(r-1)\{F^{-1}(p_r)\}^2 f_Y(p_r)},$$

so that

$$\text{Var}\{\phi(U_r)\} \doteq \left\{ \frac{1}{(r-1)\{F^{-1}(p_r)\}^2 f_Y(p_r)} \right\}^2 \text{Var}(U_r).$$

Now, using (2.16) and then making the appropriate substitutions for r , p_r and $\phi(\cdot)$, we get

$$\begin{aligned} \text{Var}\{\hat{f}_Y(0)\} &= \frac{1}{N \left[F^{-1} \left(\frac{\sqrt{N}+1}{N+1} \right) \right]^4 \left[f_Y \left(\frac{\sqrt{N}+1}{N+1} \right) \right]^2} \times \frac{(\sqrt{N}+1)(N-\sqrt{N})}{(N+1)^2(N+2)} \\ &= \frac{1}{\left[F^{-1} \left(\frac{\sqrt{N}+1}{N+1} \right) \right]^4 \left[f_Y \left(\frac{\sqrt{N}+1}{N+1} \right) \right]^2} \times \frac{(\sqrt{N}+1)(N-\sqrt{N})N}{(N+1)^2(N+2)} \cdot \frac{1}{\sqrt{N}}. \end{aligned}$$

Therefore, as $N \rightarrow \infty$ we have

$$\lim_{N \rightarrow \infty} \sqrt{N} \{\text{Var} \hat{f}_Y(0)\} = f_Y^2(0),$$

so that an approximation for the variance, when N is large, is given by

$$\text{Var}\{\hat{f}_Y(0)\} = \frac{f_Y^2(0)}{\sqrt{N}}. \quad (2.19)$$

As stated earlier, the expressions obtained for the expected value and variance of $\hat{f}_Y(0)$ are only approximations. Their adequacy for practical purposes may be evaluated by a Monte Carlo study involving various specific forms for the p.d.f., $f_Y(\cdot)$. In the next section we will look at the results of just this kind of simulation study.

2.4.4 A Monte Carlo Study

A Monte Carlo study was used to examine the approximations for $E\{\hat{f}_Y(0)\}$ and $\text{Var}\{\hat{f}_Y(0)\}$ presented in Section 2.4.3. Three possible shapes for $f_Y(\cdot)$ were used in the study. Since the shape of $f_Y(\cdot)$ depends solely on the choice of $g(y)$, the functions

$$g_1(y) = e^{-10y}, \quad y \geq 0$$

$$g_2(y) = 1-y, \quad 0 < y < 1$$

and

$$g_3(y) = 1 - y^2, \quad 0 < y < 1$$

were chosen. The function $g_1(\cdot)$ was first proposed by Gates et al. (1968), while Eberhardt (1968) suggested both $g_2(\cdot)$ and $g_3(\cdot)$. The different shapes that these three functions represent are depicted in Figure 6.

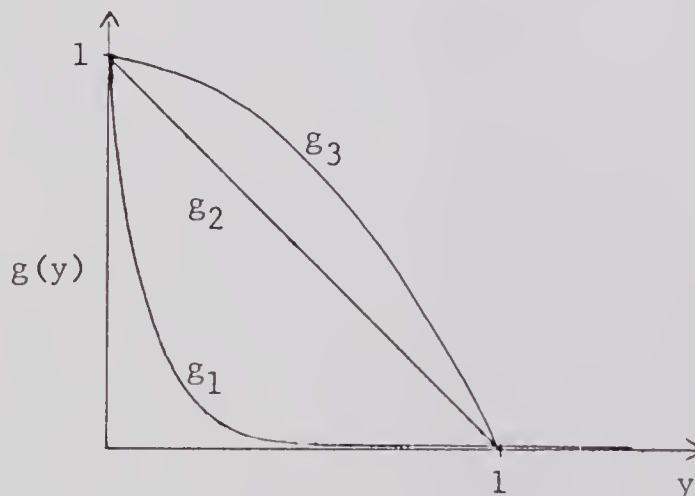


Figure 6. Three forms for the function $g(y)$.

For each value of $n = 25, 35, 45, 65, 80$ and 100 , two thousand random samples of size n were selected from each of the three populations defined by $g_i(\cdot)$, $i = 1, 2, 3$. These samples were obtained by first generating observations from a uniform distribution defined on the interval $[0, 1]$ and then transforming these values using the appropriate density $f_Y(\cdot)$. The UNIFORM function described in Barr et al. (1976) was used to generate the samples from the uniform distribution. For each set of 2000 samples, empirical estimates were calculated for the expected value, μ_e , the percent bias, B_e , and the

standard deviation, σ_e , of $\hat{f}_Y(0)$ given in equation (2.15) as follows:

Let $f_{iY}(0)$ denote the estimate from the i^{th} sample, $i = 1, 2, \dots, 2000$. Then

$$\mu_e = \frac{1}{2000} \sum_{i=1}^{2000} \hat{f}_{iY}(0)$$

$$B_e = 100 \left(\frac{\mu_e - f_Y(0)}{f_Y(0)} \right)$$

and

$$\sigma_e = \frac{1}{1999} \left(\sum_{i=1}^{2000} (\hat{f}_{iY}(0) - \mu_e)^2 \right)^{1/2}.$$

All of the necessary computing was performed under release 76.6B of SAS (see Barr et al., 1976) at the Northeast Regional Data Center located at the University of Florida.

The results of the study, along with the approximate standard deviations,

$$\sigma_T = \frac{f_Y(0)}{\sqrt{N}},$$

are presented in Tables 3, 4, and 5. As can be seen from the tables, the estimate of $f_Y(0)$ has a negative bias for most samples, generally of a magnitude less than 10% of the true value.

The ratio of σ_T/σ_e is also within 10% of one for almost all samples considered. This is even true for the smaller sample sizes, $n \leq 45$. Also, when considering the smaller sample sizes, the ratio was for the most part greater than one. Based on the results of this simulation, we feel that, in practice, the approximations obtained for the expected value and variance of $\hat{f}_Y(0)$ would perform adequately.

Table 3. Results of Monte Carlo Study using $g_1(y) = e^{-10y}$.

Sample Size	μ_e	B_e^*	σ_T	σ_e	$\frac{\sigma_T}{\sigma_e}$
25	9.05	-9.5	4.47	4.55	.98
35	9.08	-9.2	4.11	4.12	1.00
45	8.87	-11.3	3.86	3.65	1.06
65	9.48	-5.2	3.52	3.65	.96
80	9.49	-5.1	3.35	3.56	.94
100	9.48	-5.2	3.16	3.13	1.01

*For $g_1(y)$ the theoretical mean is 10.

Table 4. Results of Monte Carlo Study using $g_2(y) = 1-y$.

Sample Size	μ_e	B_e^*	σ_T	σ_e	$\frac{\sigma_T}{\sigma_e}$
25	1.88	-6.0	.894	.850	1.05
35	1.88	-6.0	.822	.801	1.03
45	1.83	-8.5	.772	.674	1.15
65	1.96	-2.0	.704	.722	.98
80	1.92	-4.0	.669	.647	1.03
100	1.93	-3.5	.632	.615	1.03

*For $g_2(y)$ the theoretical mean is 2.

Table 5. Results of Monte Carlo Study using $g_3(y) = 1-y^2$.

Sample Size	μ_e	B_e^*	σ_T	σ_e	$\frac{\sigma_T}{\sigma_e}$
25	1.47	-2.0	.671	.625	1.07
35	1.48	-1.3	.616	.594	1.04
45	1.44	-4.0	.579	.559	1.04
65	1.51	.7	.528	.536	.99
80	1.49	- .7	.502	.506	.99
100	1.50	0.0	.474	.477	.99

*For $g_3(y)$ the theoretical mean is 1.5.

2.4.5 The Expected Value and Variance for a Nonparametric Estimate of D

Now that we have decided upon an estimate for $f_Y(0)$, the problem of estimating D is straightforward. Substituting the estimate, $\hat{f}_Y(0)$, defined in Section 2.4.2 into expression (2.14) a nonparametric estimate for D is

$$\hat{D}_N = \frac{(N_o - 1)}{2L\sqrt{N_o} Y_{([\sqrt{N_o} + 1])}} \quad (2.20)$$

Expressions for the expected value and variance of $\hat{D}_N$ are easily obtained. Since L and $Y_1, Y_2, \dots, Y_{N_o}$ are independent, we can write

$$E(\hat{D}_N) = \frac{1}{2} E(\hat{\theta}) E\{\hat{f}_Y(0)\},$$

and (see Goodman, 1960, Eq. (2))

$$\text{Var}(\hat{D}_N) = \frac{1}{4} \left[E^2(\hat{\theta}) \text{Var}\{\hat{f}_Y(0)\} + E^2\{\hat{f}_Y(0)\} \text{Var}(\hat{\theta}) + \text{Var}(\hat{\theta}) \text{Var}\{\hat{f}_Y(0)\} \right],$$

where

$$\hat{\theta} = \frac{N_o - 1}{L},$$

and

$$\hat{f}_Y(0) = \frac{1}{\sqrt{N_o} Y([\sqrt{N_o} + 1])}.$$

Then, upon substituting the appropriate expressions for the moments of $\hat{\theta}$ and $\hat{f}_Y(0)$ into the above equations, we get

$$E(\hat{D}_N) \doteq D, \quad (2.21)$$

and

$$\text{Var}(\hat{D}_N) \doteq D^2 \frac{(\sqrt{N_o} + 1)}{(N_o + 2)}. \quad (2.22)$$

2.4.6 Sample Size Determination Using $\hat{D}_N$

We can now determine the approximate value of N_o that is needed to guarantee some preset value for the coefficient of variation of $\hat{D}_N$, $CV(\hat{D}_N)$. These values for N_o can then be compared to the corresponding values for N_o (see Table 1) that are needed to ensure the same coefficient of variation with the parametric estimate, $\hat{D}_u$. Using (2.21) and (2.22), we see that an approximation for the coefficient of variation of $\hat{D}_N$ is

$$CV(\hat{D}_N) \doteq \frac{(\sqrt{N_o} + 1)^{1/2}}{(N_o + 2)^{1/2}},$$

and by setting $C = CV(\hat{D}_N)$, one can easily show that $\sqrt{N_o}$ is the root of the quadratic equation

$$C^2 N_o - \sqrt{N_o} + 2C^2 - 1 = 0.$$

Solving for $\sqrt{N_0}$ yields the two roots

$$\sqrt{N_0} = \frac{1 \pm (1 - 4C^2(2C^2 - 1))^{1/2}}{2C^2}$$

and since

$$(1 - 4C^2(2C^2 - 1))^{1/2} \geq 1$$

whenever

$$C^2 < \frac{1}{2},$$

the required sample size for values of $C \leq .5$ is

$$N_0 = \left\{ \frac{1 + (1 - 4C^2(2C^2 - 1))^{1/2}}{2C^2} \right\}^2.$$

For example, if $C = .25$, then $N_0 = 284$. Table 6 gives values for N_0 corresponding to coefficients of variation ranging from .2 to .5.

Table 6. Number of animals, N_0 , that must be sighted to guarantee the estimate $\hat{D}_N$ has coefficient of variation, $CV(\hat{D}_N)$.

$CV(\hat{D}_N)$	N_0
.50	20
.40	48
.30	142
.25	284
.20	671

CHAPTER III

DENSITY ESTIMATION BASED ON A COMBINATION OF INVERSE AND DIRECT SAMPLING

3.1 Introduction

When sampling a population by means of line transects, it is important to keep in mind that the transect length that can be covered by an observer will be finite. This poses a problem for the inverse sampling plan since there will exist the possibility of not seeing the specified number of animals within the entire length of the transect. Therefore, it seems reasonable to develop a sampling scheme that would employ a rule, which allows one to stop when either a specified number, N_0 , of animals are seen or a fixed distance, L_0 , has been travelled on the transect. In this chapter we will consider a sampling plan which combines the inverse sampling procedure discussed in Chapter II and the direct sampling procedure of Gates et al. (1968).

More precisely, we will define the combined sampling method as follows:

1. Place a line at random across the area, A , to be sampled
2. Specify a fixed number of animals, $N_0 > 2$, and a fixed transect length, L_0 , and then continue sampling along the transect until either N_0 animals are seen or a distance, L_0 , has been travelled.

Since the above method merely incorporates the individual stopping rules from the inverse and direct sampling methods, it seems reasonable to use the estimate

$$\hat{D}_{CP} = \begin{cases} \hat{D}_u & \text{if } N = N_0 \\ \hat{D}_g & \text{if } N < N_0, \end{cases} \quad (3.1)$$

where N is a random variable corresponding to the actual number of animals sighted using combined sampling, $\hat{D}_u$ is the inverse sampling estimator given in (2.7) and $\hat{D}_g$ is an estimator appropriate for the direct sampling case. In other words, the combined sampling procedure uses the inverse sampling estimate if sampling terminates after N_0 animals are seen and the direct sampling estimate if sampling terminates after travelling a distance L_0 . In Section 3.5 we will also show that D_{CP} has a maximum likelihood justification.

Before proceeding to derive the mean and variance for $\hat{D}_{CP}$, we need an estimate appropriate for the direct sampling case.

3.2 Gates Estimate

Based on the direct sampling approach and assuming

$$g(y) = e^{-\lambda y}, \quad \lambda > 0,$$

Gates et al. (1968) developed the estimate

$$\hat{D}_d = \begin{cases} 0 & , n = 0, 1 \\ \frac{n(n-1)}{2L_0 \sum_{i=1}^n y_i} & , n \geq 2. \end{cases} \quad (3.2)$$

where L_0 is the fixed length of the transect, n is an observed value for the random variable N_d , the number of animals seen using direct sampling, and y_i is an observed value for the random variable Y_i , $i=1,2,\dots,n$, the right angle distance to the i^{th} animal seen. In what follows, we shall show that the variance of D_d is not finite. First, we need a result concerning the joint density of the Y_i , $i=1,2,\dots,N_d$, conditional on N_d .

Theorem 3. Under the assumptions stated in Section 2.2.1, conditional on $N_d = n > 0$, the random variables $Y_1, Y_2, \dots, Y_{N_d}$ are independently, identically distributed with common density

$$f_Y(y) = \lambda e^{-\lambda y} \quad y > 0, \quad \lambda > 0.$$

Consequently, conditional on $N_d = n > 0$, the random variable $\sum_{i=1}^{N_d} Y_i$ has a Gamma distribution with parameters n and λ .

Proof: We want to show that for $y_i > 0$, $i=1,2,\dots,N_d$,

$$f_{Y_1 Y_2 \dots Y_{N_d}}(y_1, y_2, \dots, y_{N_d} | N_d = n) = \lambda^n e^{-\lambda \sum_{i=1}^n y_i}.$$

Recall that in the direct sampling procedure, the total length travelled, L_0 , is fixed, and define L_n to be the random variable representing the total length travelled on the transect when the n^{th} animal is sighted. Then the events $\{N_d = n\}$ and $\{L_n \leq L_0 < L_{n+1}\}$, are equivalent so that

$$\begin{aligned}
f_{Y_1 Y_2 \dots Y_{N_d}}(y_1, y_2, \dots, y_{N_d} | N_d = n) \\
= f_{Y_1 Y_2 \dots Y_n}(y_1, y_2, \dots, y_n | L_n \leq L_0 < L_{n+1}).
\end{aligned}$$

Now by Theorem 2, $Y_1, Y_2, \dots, Y_n, L_n$ and L_{n+1} are mutually independent, and

$$f_{Y_i}(y_i) = \frac{g(y_i)}{c}.$$

Consequently,

$$\begin{aligned}
f_{Y_1 Y_2 \dots Y_{N_d}}(y_1, y_2, \dots, y_{N_d} | N_d = n) &= \prod_{i=1}^n f_{Y_i}(y_i) \\
&= \frac{1}{c^n} \prod_{i=1}^n g(y_i). \\
&= \lambda^n \prod_{i=1}^n e^{-\lambda y_i},
\end{aligned}$$

which completes the proof.

It is now easy to show that $\text{Var}(\hat{D}_d)$ does not exist.

From Theorem 3, conditional on $N_d = n > 0$, $\sum_{i=1}^{N_d} Y_i$ has a Gamma distribution with parameters n and λ . Thus, using (2.12) and (3.2)

$$E(\hat{D}_d^2 | N_d = n) = \begin{cases} 0 & , \quad n = 0, 1 \\ \frac{n^2(n-1)\lambda^2}{4L_0^2(n-2)} & , \quad n \geq 2 \end{cases}.$$

Also, since N_d is the number of sightings in a transect of length L_o , it follows from Theorem 2 that N_d has a Poisson distribution with parameter θL_o . Thus

$$\begin{aligned} E(\hat{D}^2) &= E_{N_d} E(\hat{D}^2 | N_d) \\ &= \frac{\lambda^2}{2} \sum_{n=2}^{\infty} \frac{n^2(n-1)}{(n-2)!} \frac{e^{-\theta L_o} (\theta L_o)^n}{n!} \\ &= +\infty, \end{aligned}$$

showing that the variance for the estimate $\hat{D}_d$ defined in (3.2) is infinite. In fact as long as

$$P(N_d = 2) > 0,$$

the variance of $\hat{D}_d$ cannot be finite.

The problem of infinite variance for $\hat{D}_d$ can be overcome by replacing $\hat{D}_d$ with $\hat{D}_g$, where

$$\hat{D}_g = \begin{cases} 0, & \text{if } n=0,1,2 \\ \frac{n(n-1)}{2L_o \sum_{i=1}^n Y_i}, & \text{if } n \geq 3 \end{cases}. \quad (3.3)$$

Note that the estimate, $\hat{D}_g$, differs from $\hat{D}_d$ only when $n=2$. Since any estimate of the density based on only 2 sightings should be effectively 0, the above modification does not seem to be unreasonable. We will now proceed to derive expressions for the mean and variance of $\hat{D}_g$, which are needed in the sequel.

3.2.1 The Mean and Variance of $\hat{D}_g$

We will first examine $E(\hat{D}_g)$. Recall from Theorem 3, that conditional on $N_d=n$, $n>0$, $\sum_{i=1}^{N_d} Y_i$ has a Gamma distribution with parameters n and λ . Thus

$$E\left(\frac{1}{\sum_{i=1}^{N_d} Y_i} \mid N_d=n\right) = \frac{\lambda}{n-1}$$

and

$$E(\hat{D}_g \mid N_d=n) = \begin{cases} 0 & , \quad n=0,1,2 \\ \frac{n\lambda}{2L_o} & , \quad n \geq 3 \end{cases}.$$

Now since N_d is distributed as a Poisson random variable with parameter θL_o , it follows that

$$\begin{aligned} E(\hat{D}_g) &= E_{N_d} E(\hat{D}_g \mid N_d) \\ &= \frac{\lambda}{2L_o} \sum_{n=3}^{\infty} \frac{ne^{-\theta L_o} (\theta L_o)^n}{n!} \\ &= \frac{\theta \lambda}{2} \{1 - e^{-\theta L_o} (1 + \theta L_o)\}. \end{aligned}$$

Substituting the left hand side of (2.5) for $\frac{\theta \lambda}{2}$ in the above yields

$$E(\hat{D}_g) = D\{1 - e^{-\theta L_o} (1 + \theta L_o)\},$$

and after writing $\mu = \theta L_o$, the expected number of sightings in a transect of length L_o , we get

$$E(\hat{D}_g) = D\{1 - e^{-\mu} (1 + \mu)\}. \quad (3.4)$$

Thus, $\hat{D}_g$ is not strictly unbiased, but the bias arises because there is a positive probability of obtaining samples of size 1 or 2. However, even for moderate values of μ , the bias in $\hat{D}_g$ will be small since $e^{-\mu}(1+\mu)$ tends to zero exponentially fast. For example, if $\mu = 10$, the relative bias is only .05%.

Next we will look at $\text{Var}(\hat{D}_g)$. Again since, conditional on $N_d=n$, $n>0$, $\sum_{i=1}^{N_d} Y_i$ is distributed as a Gamma random variable, with parameters n and λ , we know that

$$E \left[\frac{1}{\left(\sum_{i=1}^{N_d} Y_i \right)^2} \middle| N_d=n \right] = \frac{\lambda^2}{(n-1)(n-2)}, \quad n>2,$$

and

$$E(\hat{D}_g^2 | N_d=n) = \begin{cases} 0 & , \text{ if } n=0,1,2 \\ \frac{n^2(n-1)\lambda^2}{4L_o^2(n-2)} & , \text{ if } n \geq 3 \end{cases}.$$

Therefore,

$$\begin{aligned} E(\hat{D}_g^2) &= E_{N_d} E(\hat{D}_g^2 | N_d) \\ &= \frac{\lambda^2}{4L_o^2} \sum_{n=3}^{\infty} \frac{n^2(n-1)}{(n-2)} \frac{e^{-\mu} \mu^n}{n!} \end{aligned} \quad (3.5)$$

and we can write

$$\text{Var}(\hat{D}_g) = \frac{\lambda^2}{4L_o^2} \sum_{n=3}^{\infty} \frac{n^2(n-1)}{(n-2)} \frac{e^{-\mu} \mu^n}{n!} - D^2 \{1 - e^{-\mu}(1+\mu)\}^2. \quad (3.6)$$

An approximation to $\text{Var}(\hat{D}_g)$ valid for large values of μ may be derived in a manner analogous to the method used by Gates et al. (1968). After writing

$$\frac{n^2(n-1)}{(n-2)} = n^{2+n+2+\frac{4}{(n-2)}},$$

it is easy to see that for $n \geq 3$

$$n^{2+n+2} \leq \frac{n^2(n-1)}{(n-2)} \leq n^{2+n+6}.$$

Thus, lower and upper bounds for $E(D_g^2)$ are

$$LB = \frac{\lambda^2}{4L_o^2} \sum_{n=3}^{\infty} (n^{2+n+2}) \frac{e^{-\mu} \mu^n}{n!} \leq E(\hat{D}_g^2)$$

and

$$UB = \frac{\lambda^2}{4L_o^2} \sum_{n=3}^{\infty} (n^{2+n+6}) \frac{e^{-\mu} \mu^n}{n!} \geq E(\hat{D}_g^2).$$

Now

$$\begin{aligned} UB-LB &= \frac{\lambda^2}{L_o^2} \sum_{n=3}^{\infty} \frac{e^{-\mu} \mu^n}{n!} \\ &= \frac{\lambda^2}{L_o^2} (1 - e^{-\mu} - \mu e^{-\mu} - \frac{\mu^2 e^{-\mu}}{2}). \end{aligned}$$

Upon using the relationships

$$D = \frac{\lambda \theta}{2} \quad \text{and} \quad \mu = \theta L_o$$

we get

$$UB-LB = \frac{4D^2}{\mu^2} (1 - e^{-\mu} - \mu e^{-\mu} - \frac{\mu^2 e^{-\mu}}{2}),$$

which tends to 0 as $\mu \rightarrow \infty$. Thus, a reasonable approximation for $E(\hat{D}_g^2)$ is

$$\begin{aligned} E(\hat{D}_g^2) &\doteq \frac{UB+LB}{2} \\ &= \frac{\lambda^2}{4L_o^2} \sum_{n=3}^{\infty} (n^{2+n+4}) \frac{e^{-\mu} \mu^n}{n!} \\ &= \frac{D^2}{\mu^2} \{ \mu^{2+2\mu+4} - e^{-\mu} (4+6\mu+5\mu^2) \} \end{aligned} \quad (3.7)$$

From (3.7) an approximation for $\text{Var}(\hat{D}_g)$ is

$$\begin{aligned}\text{Var}(\hat{D}_g) &= D^2 \left\{ 1 + \frac{2}{\mu} + \frac{4}{\mu^2} - e^{-\mu} \left(5 + \frac{6}{\mu} + \frac{4}{\mu^2} \right) \right\} - D^2 \{ 1 - 2e^{-\mu}(1+\mu) + e^{-2\mu}(1+\mu)^2 \} \\ &= D^2 \left\{ \frac{2}{\mu} + \frac{4}{\mu^2} - e^{-\mu} \left(3 - 2\mu + \frac{6}{\mu} + \frac{4}{\mu^2} \right) - e^{-2\mu}(1+\mu)^2 \right\} .\end{aligned}$$

Now, as μ increases, the terms involving $e^{-\mu}$ and $e^{-2\mu}$ will tend to 0 much faster than $\frac{2}{\mu} + \frac{4}{\mu^2}$, so that for large μ , we have the approximation

$$\text{Var}(\hat{D}_g) \doteq D^2 \left(\frac{2}{\mu} + \frac{4}{\mu^2} \right). \quad (3.8)$$

We are now in a position to derive the mean and variance of $\hat{D}_{CP}$.

3.3 Expected Value of $\hat{D}_{CP}$

Recall that in the combined sampling scheme both N , the number of animals seen, and L , the distance travelled before termination of sampling are random variables. Thus, the expected value of D_{CP} can be found directly using

$$E(\hat{D}_{CP}) = E_N E(\hat{D}_{CP} | N).$$

However, before proceeding along these lines it will be helpful to have the following theorems.

Theorem 4. Let N be the random variable representing the number of animals seen using the combined sampling method. Then under the assumptions stated in Section 2.2.1,

$$P(N=n) = \begin{cases} \frac{e^{-\mu} \mu^n}{n!} & , n=0,1,\dots,N_0-1 \\ \sum_{j=N_0}^{\infty} \frac{e^{-\mu} \mu^j}{j!} & , n=N_0 \end{cases}$$

where $\mu = \theta L_0$ is the expected number of animals sighted along a transect of length L_0 .

Proof: For $n < N_0$, the event $\{N=n\}$ is equivalent to the event $\{\text{exactly } n \text{ sightings occur in } (0, L_0]\}$. By Theorem 1, the number of sightings in $(0, L_0]$ is Poisson with parameter θL_0 . Hence,

$$P(N=n) = \frac{e^{-\mu} \mu^n}{n!} , n=0, \dots, N_0-1.$$

The case $N=N_0$ follows since the event $\{N=N_0\}$ is equivalent to the event $\{\text{at least } N_0 \text{ sightings occur in } (0, L_0]\}$.

The following three theorems establish some useful relationships among the random variables N , L_N and $Y_1, Y_2, \dots, Y_N$ where N is as defined in Theorem 4, L_N represents the total length travelled on the transect when the N^{th} animal is sighted and Y_i , $i=1, 2, \dots, N$, represents the right angle distance to the i^{th} animal seen.

Theorem 5. Under the assumptions stated in Section 2.2.1, the conditional p.d.f. of L_N given $N=n>0$, is

$$f_{L_N}(\ell | N=n) = \begin{cases} \frac{n\ell^{n-1}}{L_o^n} & , \quad \ell < L_o, \quad n=1,2,\dots,N_o-1 \\ \frac{\theta^{N_o} \ell^{N_o-1} e^{-\theta\ell}}{\Gamma(N_o)} \cdot \frac{1}{P(N=N_o)} & \ell < L_o, \quad n=N_o \end{cases}$$

where

$$P(N=N_o) = \int_0^{L_o} \frac{\theta^{N_o} \ell^{N_o-1} e^{-\theta\ell}}{\Gamma(N_o)} d\ell.$$

Proof: First we will consider the case when $n < N_o$. By Theorem 1, seeing $n < N_o$ animals in distance L_o is equivalent to observing n Poisson events in the interval $(0, L_o]$. Therefore (see Bhat, 1972, p. 129), the joint density of $L_1, L_2, \dots, L_N$ conditional on the occurrence of $N=n$ Poisson events in $(0, L_o]$ is

$$f_{L_1, L_2, \dots, L_N}(\ell_1, \ell_2, \dots, \ell_n | N=n) = \frac{n!}{L_o^n}, \quad 0 \leq \ell_1 \leq \ell_2 \leq \dots \leq \ell_n \leq L_o,$$

and the marginal density of L_N conditional on $N=n$ is

$$f_{L_N}(\ell | N=n) = n! \frac{\ell^{n-1}}{L_o^n}, \quad 0 < \ell < L_o.$$

Next, we will consider the case where $N=N_o$. Define T_i to be the random variable corresponding to the distance travelled on the transect between the $(i-1)^{st}$ and i^{th} sighting. Then the N_o^{th} observation is made at

$$L_{N_o} = \sum_{i=1}^{N_o} T_i.$$

Now in the combined sampling approach, we will see $N=N_0$ animals, if and only if the distance

$$L_{N_0} = \sum_{i=1}^{N_0} T_i \leq L_0.$$

Therefore, if $0 \leq \ell \leq L_0$, then

$$\begin{aligned} P(L_{N_0} \leq \ell | N=N_0) &= \frac{P(L_{N_0} \leq \ell, N=N_0)}{P(N=N_0)} \\ &= \frac{P(L_{N_0} \leq \ell, L_{N_0} \leq L_0)}{P(N=N_0)} \\ &= \frac{P(\sum_{i=1}^{N_0} T_i \leq \ell)}{P(N=N_0)}. \end{aligned}$$

Now, since the sightings are Poisson events by Theorem 1, the random variable

$$L_{N_0} = \sum_{i=1}^{N_0} T_i$$

has a Gamma distribution with parameters N_0 and θ . Thus,

$$f_{L_{N_0}}(\ell | N=N_0) = \frac{\theta^{N_0} \ell^{N_0-1} e^{-\theta \ell}}{\Gamma(N_0)} \frac{1}{P(N=N_0)}, \quad 0 \leq \ell \leq L_0. \quad (3.9)$$

Now, by Theorem 4, we have

$$\begin{aligned} P(N=N_0) &= \sum_{j=N_0}^{\infty} e^{-\theta L_0} \frac{(\theta L_0)^j}{j!} \\ &= \int_0^{\theta L_0} \frac{e^{-z} z^{N_0-1}}{\Gamma(N_0)} dz \\ &= \int_0^{L_0} \frac{e^{-\theta \ell} \theta^{N_0} \ell^{N_0-1}}{\Gamma(N_0)} d\ell. \end{aligned}$$

Substituting $P(N=N_0)$ into (3.9) above completes the proof.

Theorem 6. Under the assumptions stated in Section 2.2.1, conditional on $N=n>0$, the random variables $Y_1, Y_2, \dots, Y_N$ and L_N are independent.

Proof: First consider the case $N=N_0$. Let $\ell \geq 0$ and $y_i \geq 0$ for $i=1, 2, \dots, N$. We want to show that

$$\begin{aligned} & P(Y_i \leq y_i, L_N \leq \ell, i = 1, 2, \dots, N | N=N_0) \\ &= P(Y_i \leq y_i, i=1, 2, \dots, N_0) P(L_{N_0} \leq \ell | N=N_0). \end{aligned}$$

Note that the event $\{N=N_0\}$ is equivalent to the event $\{L_{N_0} \leq L_0\}$, so that we can write

$$\begin{aligned} & P(Y_i \leq y_i, L_N \leq \ell, i = 1, \dots, N | N=N_0) \\ &= P(Y_i \leq y_i, L_{N_0} \leq \ell, i = 1, 2, \dots, N_0 | N=N_0) \\ &= P(Y_i \leq y_i, L_{N_0} \leq \ell, i = 1, 2, \dots, N_0 | L_{N_0} \leq L_0) \\ &= \frac{P(Y_i \leq y_i, L_{N_0} \leq \ell, L_{N_0} \leq L_0, i=1, 2, \dots, N_0)}{P(L_{N_0} \leq L_0)}. \end{aligned}$$

Now by Theorem 2, $Y_1, Y_2, \dots, Y_{N_0}$ and L_{N_0} are independent.

Consequently,

$$\begin{aligned} & \frac{P(Y_i \leq y_i, L_{N_0} \leq \ell, L_{N_0} \leq L_0, i=1, 2, \dots, N_0)}{P(L_{N_0} \leq L_0)} \\ &= \frac{P(Y_i \leq y_i, i = 1, 2, \dots, N_0) P(L_{N_0} \leq \ell, L_{N_0} \leq L_0)}{P(L_{N_0} \leq L_0)} \\ &= P(Y_i \leq y_i, i=1, 2, \dots, N_0) P(L_{N_0} \leq \ell | L_{N_0} \leq L_0) \\ &= P(Y_i \leq y_i, i=1, 2, \dots, N_0) P(L_{N_0} \leq \ell | N=N_0). \end{aligned}$$

Now consider the case $N=n < N_0$, and let ℓ and y_i be defined as before. Also, define X_N to be the actual length travelled to see N animals when the combined sampling method is used, that is,

$$X_N = \begin{cases} L_0, & N < N_0 \\ L_{N_0}, & N = N_0 \end{cases}.$$

Then for $n < N_0$, the events $\{N=n\}$, $\{X_N=L_0\}$ and $\{L_n \leq L_0 < L_{n+1}\}$ are equivalent. Thus,

$$\begin{aligned} P(Y_i \leq y_i, L_n \leq \ell, i = 1, 2, \dots, n | N=n < N_0) \\ &= P(Y_i \leq y_i, L_n \leq \ell, i = 1, 2, \dots, n | X_N=L_0) \\ &= P(Y_i \leq y_i, L_n \leq \ell, i = 1, 2, \dots, n | L_n \leq L_0 < L_{n+1}) \\ &= \frac{P(Y_i \leq y_i, L_n \leq \ell, L_n \leq L_0 < L_{n+1}, i=1, 2, \dots, n)}{P(L_n \leq L_0 < L_{n+1})}. \end{aligned}$$

Again by Theorem 2, for $N=n$, $Y_1, Y_2, \dots, Y_n$ and L_n, L_{n+1} , are independent, so that

$$\begin{aligned} &\frac{P(Y_i \leq y_i, L_n \leq \ell, L_n \leq L_0 < L_{n+1}, i = 1, 2, \dots, n)}{P(L_n \leq L_0 < L_{n+1})} \\ &= \frac{P(Y_i \leq y_i, i=1, 2, \dots, n) P(L_n \leq \ell, L_n \leq L_0 < L_{n+1})}{P(L_n \leq L_0 < L_{n+1})} \\ &= P(Y_i \leq y_i, i=1, 2, \dots, n) P(L_n \leq \ell | N=n < N_0). \end{aligned}$$

Theorem 7. Under the assumptions stated in Section 2.2.1, conditional on $N=n > 0$, the random variables $Y_1, Y_2, \dots, Y_N$ are independently, identically distributed with common density

$$f_Y(y) = \lambda e^{-\lambda y}, \quad y > 0, \lambda > 0.$$

Consequently, conditional on $N=n>0$, the random variable $\sum_{i=1}^N Y_i$ has a Gamma distribution with parameters n and λ .

Proof: The case $N=n < N_0$ follows directly from Theorem 3 by noting that the random variables N and N_d are equivalent when $0 < n < N_0$. Now consider the case $N=N_0$. The event $\{N=N_0\}$ is equivalent to the event $\{L_{N_0} \leq L_0\}$ so that we can write

$$\begin{aligned} f_{Y_1, Y_2, \dots, Y_N}(y_1, y_2, \dots, y_N | N=N_0) \\ = f_{Y_1, Y_2, \dots, Y_{N_0}}(y_1, y_2, \dots, y_{N_0} | L_{N_0} \leq L_0). \end{aligned}$$

Now by Theorem 2, $Y_1, Y_2, \dots, Y_{N_0}$ and L_{N_0} are mutually independent and

$$f_{Y_i}(y_i) = \frac{g(y_i)}{c}.$$

Consequently,

$$\begin{aligned} f_{Y_1 Y_2 \dots Y_{N_0}}(y_1, y_2, \dots, y_{N_0} | N=N_0) &= \prod_{i=1}^{N_0} f_{Y_i}(y_i) \\ &= \frac{1}{c^{N_0}} \prod_{i=1}^{N_0} g(y_i), \end{aligned}$$

and substituting

$$g(y_i) = e^{-\lambda y_i}$$

completes the proof.

We are now ready to determine the expected value of $\hat{D}_{CP}$ given $N=n$. For $n = 0, 1, 2$,

$$E(\hat{D}_{CP} | N=n) = E(\hat{D}_g | N=n) = 0. \quad (3.10)$$

Next consider the values $3 \leq n < N_0$. Recall from Theorem 7 that, conditional on $N=n>0$, $\sum_{i=1}^N Y_i$ has a Gamma distribution with parameters n and λ . Then using expressions (3.1) and (3.3) it follows that

$$\begin{aligned} E(\hat{D}_{CP} | N=n) &= E(\hat{D}_g | N=n) \\ &= E \left(\frac{N(N-1)}{2L_0 \sum_{i=1}^N Y_i} \mid N=n \right) \\ &= \frac{n\lambda}{2L_0}. \end{aligned} \quad (3.11)$$

Finally for $N=N_0$ it follows from Theorem 5, Theorem 6, Theorem 7, and expressions (3.1) and (3.3) that

$$\begin{aligned} E(\hat{D}_{CP} | N=N_0) &= E(\hat{D}_u | N=N_0) \\ &= \frac{1}{2} E \left[\frac{N-1}{\sum_{i=1}^N Y_i} \mid N=N_0 \right] E \left[\frac{N-1}{L_N} \mid N=N_0 \right] \\ &= \frac{\lambda \theta}{2P(N=N_0)} \cdot \int_0^{L_0} \frac{\theta^{N_0-1} \ell^{N_0-2} e^{-\theta \ell}}{\Gamma(N_0-1)} d\ell. \end{aligned}$$

Then using the transformation $\ell = \frac{Z}{\theta}$ we get

$$E(\hat{D}_{CP} | N=N_0) = \frac{\lambda \theta}{2P(N=N_0)} \sum_{n=N_0-1}^{\infty} \frac{e^{-\theta L_0} (\theta L_0)^n}{n!}. \quad (3.12)$$

We can now evaluate the expected value of $\hat{D}_{CP}$. Using Theorem 4 and expressions (2.5), (3.10), (3.11), and (3.12), we find that

$$\begin{aligned} E(\hat{D}_{CP}) &= E_N E(\hat{D}_{CP} | N) \\ &= \sum_{n=3}^{N_0-1} \frac{n\lambda}{2L_0} P(N=n) + \frac{\theta\lambda}{2} \sum_{n=N_0-1}^{\infty} \frac{e^{-\theta L_0} (\theta L_0)^n}{n!} \\ &= D[1 - e^{-\mu}(1+\mu)]. \end{aligned} \quad (3.13)$$

where $\mu = \theta L_0$.

Thus $\hat{D}_{CP}$ is a biased estimate for the density. Note that the bias here is equal to the bias for the modified estimate, $\hat{D}_g$, in direct sampling. This is as expected since in the combined sampling procedure, we are simply choosing the estimate that corresponds to the reason for terminating sampling. If we stop sampling after seeing the N_0^{th} animal, then the inverse sampling estimate is used, and, likewise, if sampling stops after travelling the distance L_0 , then the direct sampling estimate is used.

3.4 Variance of $\hat{D}_{CP}$

An expression for the variance of $\hat{D}_{CP}$ can be found directly using the formula

$$\text{Var}(\hat{D}_{CP}) = E(\hat{D}_{CP}^2) - \{E(\hat{D}_{CP})\}^2. \quad (3.14)$$

In the preceding section we derived $E(\hat{D}_{CP})$ so that our problem reduces to evaluating $E(\hat{D}_{CP}^2)$. Proceeding along the same lines as in Section (3.3), we quickly find

$$E(D_{CP}^2 | N=n, n=0,1,2) = 0, \quad (3.15)$$

$$\begin{aligned} E(\hat{D}_{CP}^2 | N=n, 2 < n < N_0) &= E \left[\frac{N^2 (N-1)^2}{4L_0^2 \left(\sum_{i=1}^N Y_i \right)^2} \middle| N=n, 2 < n < N_0 \right] \\ &= \frac{n^2 (n-1) \lambda^2}{4L_0^2 (n-2)}, \end{aligned} \quad (3.16)$$

and,

$$\begin{aligned} E(\hat{D}_{CP}^2 | N=N_0) &= E \left[\frac{(N-1)^4}{4L_N^2 \sum_{i=1}^N Y_i} \middle| N=N_0 \right] \\ &= \frac{1}{4} E \left[\frac{(N-1)^2}{L_N^2} \middle| N=N_0 \right] E \left[\frac{(N-1)^2}{\sum_{i=1}^N Y_i} \middle| N=N_0 \right] \\ &= \frac{\lambda^2 \theta^2 (N_0-1)}{4(N_0-2)^2 P(N=N_0)} \int_0^{L_0} \frac{(N_0-1)^2 \theta^{N_0-2} \ell^{N_0-3} e^{-\theta \ell}}{\Gamma(N_0)} d\ell \\ &= \frac{\lambda^2 \theta^2 (N_0-1)^2}{4(N_0-2)^2 P(N=N_0)} \int_0^{L_0} \frac{\theta^{N_0-2} \ell^{N_0-3} e^{-\theta \ell}}{\Gamma(N_0-2)} d\ell \\ &= \frac{\lambda^2 \theta^2 (N_0-1)^2}{4(N_0-2)^2 P(N=N_0)} \sum_{n=N_0-2}^{\infty} \frac{e^{-\theta L_0} (\theta L_0)^n}{n!}. \end{aligned} \quad (3.17)$$

Then, using Theorem 4, expressions (2.5), (3.15), (3.16), and (3.17) and letting $\mu = \theta L_0$, it follows that

$$\begin{aligned} E(\hat{D}_{CP}^2) &= E_N E(\hat{D}_{CP}^2 | N) \\ &= \frac{\lambda^2}{4L_0^2} \sum_{n=3}^{N_0-1} \frac{n^2 (n-1)}{(n-2)} \frac{e^{-\mu} \mu^n}{n!} + \frac{\lambda^2 \theta^2}{4} \left(\frac{N_0-1}{N_0-2} \right)^2 \sum_{n=N_0-2}^{\infty} \frac{e^{-\mu} \mu^n}{n!} \\ &= \frac{\lambda^2 \theta^2}{4} \sum_{n=3}^{N_0-1} \frac{n}{(n-2)} \frac{\mu^{n-2} e^{-\mu}}{(n-2)!} + \frac{\lambda^2 \theta^2}{4} \left(\frac{N_0-1}{N_0-2} \right)^2 \sum_{n=N_0-2}^{\infty} \frac{e^{-\mu} \mu^n}{n!} \end{aligned}$$

$$= D^2 \left[\sum_{n=1}^{N_o-3} \frac{(n+2)}{n} \frac{\mu^n e^{-\mu}}{n!} + \left(\frac{N_o-1}{N_o-2} \right)^2 \sum_{n=N_o-2}^{\infty} \frac{e^{-\mu} \mu^n}{n!} \right]. \quad (3.18)$$

An expression for the variance of $\hat{D}_{CP}$ is now evident.

Using (3.13), (3.14), and (3.18), we get

$$\text{Var}(\hat{D}_{CP}) = D^2 \left[\sum_{n=1}^{N_o-3} \frac{(n+2)}{n} \frac{\mu^n e^{-\mu}}{n!} + \left(\frac{N_o-1}{N_o-2} \right)^2 \sum_{n=N_o-2}^{\infty} \frac{e^{-\mu} \mu^n}{n!} - [1 - e^{-\mu}(1+\mu)]^2 \right], \quad (3.19)$$

where $\mu = \theta L_o$.

Note that,

$$\lim_{L_o \rightarrow \infty} \text{Var}(\hat{D}_{CP}) = \left[\left(\frac{N_o-1}{N_o-2} \right)^2 - 1 \right] D^2,$$

and

$$\lim_{N_o \rightarrow \infty} \text{Var}(\hat{D}_{CP}) = D^2 \left[\sum_{n=1}^{\infty} \frac{(n+2)}{n} \frac{\mu^n e^{-\mu}}{n!} - [1 - e^{-\mu}(1+\mu)]^2 \right].$$

After some simple algebraic manipulations and using the relationship $D = \frac{\theta \lambda}{2}$, one can easily show that the limit as $L_o \rightarrow \infty$ and the limit as $N_o \rightarrow \infty$ are equal to the $\text{Var}(\hat{D}_u)$ given in (2.13) and the $\text{Var}(\hat{D}_g)$ given in (3.6), respectively. These limiting values are as expected, since letting $L_o \rightarrow \infty$ in the combined sampling approach is equivalent to using inverse sampling, while letting $N_o \rightarrow \infty$ is equivalent to using direct sampling.

We will now show that $\text{Var}(\hat{D}_{CP})$ can be expressed as a function of both $\text{Var}(\hat{D}_u)$ and $\text{Var}(\hat{D}_g)$ given in (2.13) and (3.6), respectively. Writing the equation in this form will then lead directly to an approximation for $\text{Var}(\hat{D}_{CP})$.

First note that (3.19) can be rewritten as

$$\frac{\text{Var}(\hat{D}_{\text{CP}})}{D^2} = \sum_{n=1}^{N_o-3} \frac{(n+2)}{n} \frac{\mu^n e^{-\mu}}{n!} + \left(\frac{N_o-1}{N_o-2} \right)^2 \sum_{n=N_o-2}^{\infty} \frac{e^{-\mu} \mu^n}{n!} - [1 - e^{-\mu}(1+\mu)]^2. \quad (3.20)$$

Adding and subtracting the terms $\sum_{n=N_o-2}^{\infty} \frac{(n+2)}{n} \frac{\mu^n e^{-\mu}}{n!}$ and

$\sum_{n=N_o-2}^{\infty} \frac{e^{-\mu} \mu^n}{n!}$ to the right hand side of (3.20) yields

$$\begin{aligned} \frac{\text{Var}(\hat{D}_{\text{CP}})}{D^2} &= \sum_{n=1}^{\infty} \frac{(n+2)}{n} \frac{\mu^n e^{-\mu}}{n!} - [1 - e^{-\mu}(1+\mu)]^2 \\ &\quad + \sum_{n=N_o-2}^{\infty} \frac{e^{-\mu} \mu^n}{n!} \left[\left(\frac{N_o-1}{N_o-2} \right)^2 - 1 \right] \\ &\quad + \sum_{n=N_o-2}^{\infty} \frac{e^{-\mu} \mu^n}{n!} - \sum_{n=N_o-2}^{\infty} \frac{(n+2)}{n} \frac{\mu^n e^{-\mu}}{n!} \\ &= \sum_{n=N_o-2}^{\infty} \frac{e^{-\mu} \mu^n}{n!} \left[\frac{2N_o-3}{(N_o-2)^2} \right] + \sum_{n=1}^{\infty} \frac{(n+2)}{n} \frac{\mu^n e^{-\mu}}{n!} \\ &\quad - [1 - e^{-\mu}(1+\mu)]^2 - \sum_{n=N_o-2}^{\infty} \frac{2}{n} \frac{\mu^n e^{-\mu}}{n!}. \end{aligned}$$

Now, after multiplying both sides of the previous expression by D^2 , using the relationship $D = \frac{\theta \lambda}{2}$ and substituting the expressions for $\text{Var}(\hat{D}_u)$ and $\text{Var}(\hat{D}_g)$ given in (2.13) and (3.6), respectively, we see that

$$\text{Var}(\hat{D}_{\text{CP}}) = \left[\sum_{n=N_o-2}^{\infty} \frac{e^{-\mu} \mu^n}{n!} \right] \text{Var}(\hat{D}_u) + \text{Var}(\hat{D}_g) - D^2 \sum_{n=N_o-2}^{\infty} \frac{2}{n} \frac{\mu^n e^{-\mu}}{n!}. \quad (3.21)$$

Therefore, an approximation for $\text{Var}(\hat{D}_{\text{CP}})$ can be simply obtained by using the approximation for $\text{Var}(\hat{D}_g)$ given in (3.8).

3.5 Maximum Likelihood Justification for $\hat{D}_{CP}$

In Section 3.1 we stated that the estimate

$$\hat{D}_{CP} = \begin{cases} \hat{D}_g, & N \leq N_0 - 1 \\ \hat{D}_u, & N = N_0 \end{cases}$$

could be justified using the maximum likelihood procedure. To show this, we first need the joint density function for $Y_1, Y_2, \dots, Y_N, L_N$ and N , i.e., $f_{\underline{Y}, L_N, N}(\underline{y}, \ell, n)$.

By Theorem 6, $f_{\underline{Y}, L_N, N}(\cdot)$ can be written as

$$f_{\underline{Y}, L_N, N}(\underline{y}, \ell, n) = f_{\underline{Y}}(\underline{y} | N=n) f_{L_N}(\ell | N=n) P(N=n).$$

The functional form for $f_{\underline{Y}, L_N, N}(\cdot)$ is now evident. Using Theorems 4 and 5 and recalling from Theorem 7 that

$$f_{\underline{Y}}(\underline{y} | N=n) = \lambda^n e^{-\lambda \sum_{i=1}^n Y_i}, \quad \lambda > 0,$$

we obtain,

$$f_{\underline{Y}, L_N, N}(\underline{y}, \ell, n) = \begin{cases} \frac{\lambda^n e^{-\lambda \sum_{i=1}^n Y_i} n \ell^{n-1}}{L_0} \frac{e^{-\theta L_0} (\theta L_0)^n}{n!}, & N=n < N_0 \\ \frac{\lambda^{N_0} e^{-\lambda \sum_{i=1}^{N_0} Y_i} 0^{N_0} \ell^{N_0-1}}{(N_0-1)!} e^{-\theta \ell}, & N=N_0 \end{cases} \quad (3.22)$$

As shown in Section 2.3.1, the maximum likelihood estimate for D is given by

$$\hat{D} = \frac{\hat{\theta} \hat{\lambda}}{2},$$

where $\hat{\theta}$ and $\hat{\lambda}$ are maximum likelihood estimates for θ and λ , respectively. Finding maximum likelihood estimates for θ and λ is now straightforward. Taking the natural logarithm of the likelihood function and setting the partial derivatives with respect to θ and λ equal to 0 yields, for $n > 0$,

$$\hat{\theta} = \begin{cases} \frac{N}{L_0} & , \quad N=n < N_0 \\ \frac{N_0}{L_{N_0}} & , \quad N=N_0 \end{cases}$$

and

$$\hat{\lambda} = \frac{N}{\sum_{i=1}^N Y_i} \quad , \quad N=1, \dots, N_0.$$

Thus, a maximum likelihood estimate for D using combined sampling would be

$$\hat{D}_M = \begin{cases} \frac{N^2}{2L_0 \sum_{i=1}^N Y_i} & , \quad N=n < N_0 \\ \frac{N^2}{2L_{N_0} \sum_{i=1}^N Y_i} & , \quad N=N_0. \end{cases}$$

Our estimate for $\hat{D}_{CP}$ is obtained by correcting the estimates

$\hat{\lambda}$ and $\hat{\theta}$ for bias and noting that $E \left[\frac{1}{\left(\sum_{i=1}^N Y_i \right)^2} \mid N=n \right]$ only

exists for values of $N > 2$.

CHAPTER IV

DENSITY ESTIMATION FOR CLUSTERED POPULATIONS

4.1 Introduction

The estimation procedures developed in Chapters II and III are based on the assumption that the sightings of animals are independent events. These methods would be applicable to animal populations that are generally made up of solitary individuals, such as ruffed grouse, snowshoe hare and gila monster. However, there are other types of animals which aggregate into coveys, schools and other tight groups. Animals behaving in this way will be said to belong to clustered populations. Some examples of clustered populations are bobwhite quail, gray partridge and porpoise. In these cases the assumption of independent sightings is certainly not valid, and a different procedure would have to be used.

The line transect method could be easily generalized to provide estimates for clustered populations. As noted by Anderson et al. (1976, p. 12), if we amend the assumptions in Section 2.2.1 so that they refer to clusters of animals rather than individual animals, then the results of Chapters II and III are directly applicable to the estimation of the cluster density, D_c . The estimate for D_c will be

based on the right angle distances to the clusters from the random line transect. In the case where the number of animals in every sighted cluster can be determined without error, an estimate for the population density D is given by

$$\hat{D} = \hat{D}_c \bar{s}$$

where $\hat{D}_c$ is the estimate for D_c and $\bar{s}$ is the average size of the observed clusters.

Some criticisms of the approach outlined in the preceding paragraph are possible. First of all, it may not be possible to determine the distance to a cluster as easily (or as accurately) as the distance to an animal. How will this distance be defined? Secondly, the simple modification of the assumptions in Section 2.2.1, obtained by replacing the word "animal" by the word "cluster" would imply that the probability of sighting a cluster depends only on its right angle distance from the line. This may not be a reasonable assumption since the probability of sighting a larger cluster is likely to be greater than the probability of sighting a smaller cluster. Finally, the sighting of a cluster may not necessarily mean that all of the animals comprising the cluster are seen and counted by the observer. In this case, a more reasonable assumption would be to let the probability of sighting an animal belonging to a cluster depend on the distance to the cluster as well as the true cluster size.

In this chapter we shall propose a density estimate for a clustered population by assuming, among other things, that

it is possible to determine the distance to the center of the cluster from the line transect. An estimation procedure will then be developed using a model in which the observer's count of the number of animals in a cluster is regarded as a random variable with a probability distribution depending upon the right angle distance and the size of the cluster.

4.2 Assumptions

The density estimate that we will develop is based on the inverse sampling approach outlined in Section 2.1, with one minor modification. In clustered populations the plan is to continue sampling along a randomly placed transect until a prespecified number, N_c , of clusters (rather than animals) are seen. As each cluster is sighted, the following information is recorded:

1. the right angle distance, y , from the transect to the center of the cluster
2. the observed number of animals, s , in the cluster. (this may be less than the true size of the cluster)
3. the actual distance, ℓ , travelled by the observer to sight N_c clusters.

The sampling procedure described above may be used to construct an estimate of the population density under the following set of assumptions. These assumptions closely parallel those of Section 2.2.1 with the exception that they are now phrased in terms of clusters rather than individual animals.

- B1. The clusters are randomly distributed with rate (density) D_c over the area of interest, A .
- B2. The clusters are independently distributed over A , i.e., given two disjoint regions of area, δA_1 and δA_2 ,
- $$P(n_1 \text{ clusters are in } \delta A_1 \text{ and } n_2 \text{ clusters are in } \delta A_2) \\ = P(n_1 \text{ clusters are in } \delta A_1)P(n_2 \text{ clusters are in } \delta A_2).$$
- B3. Clusters are fixed, i.e., there is no confusion over clusters moving during sampling and none are counted twice.
- B4. There exists a probability mass function $p(\cdot)$ defined on the set of positive integers, such that $p(r)$ is the probability that r is the true size of a cluster located at a right angle distance, y , from the transect. Note that $p(r)$ is independent of y . In probability notation, if R and Y denote the random variables representing the true cluster size and the right angle distance to the cluster, respectively, then
- B5. The probability of observing a cluster depends only on the size of the cluster and the distance from the transect to the cluster.
- B6. There exists a non-negative function $h(\cdot)$ defined on $[0, \infty)$ such that

$$P(R=r|Y=y) = p(r), \quad r = 1, 2, \dots \quad (4.1)$$

$$0 \leq h(\cdot) \leq 1,$$

$$h(0) = 0,$$

and the probability of observing s animals belonging to a cluster of size $r \geq s$ located at a right angle distance y from the transect is

$$\binom{r}{s} [h(y)]^s [1-h(y)]^{r-s}.$$

That is, if Y and S denote the random variables representing the right angle distance to a cluster and the observed number of animals in a cluster, respectively, then

$$P(S=s|R=r, Y=y) = \binom{r}{s} [h(y)]^s [1-h(y)]^{r-s}. \quad (4.2)$$

Closer examination of assumption B6 shows that we are now allowing the probability of observing a cluster to depend on both the right angle distance, y , and the true cluster size, r . To see this, first let C be the event that a cluster is observed. Then the probability of observing a cluster of size r located at distance y from the transect is

$$\begin{aligned} P(C|R=r, Y=y) &= \sum_{s=1}^r P(S=s|R=r, Y=y) \\ &= 1 - P(S=0|R=r, Y=y) \\ &= 1 - [1-h(y)]^r, \end{aligned} \tag{4.3}$$

which clearly depends on both y and r .

The assumption B6 also satisfies the reasonable requirement that for a fixed right angle distance $y > 0$ and $r_1 < r_2$,

$$P(C|R=r_1, Y=y) \leq P(C|R=r_2, Y=y).$$

This follows immediately from equation (4.3). Note that

$$P(C|R=r_1, Y=y) = 1 - [1-h(y)]^{r_1},$$

and

$$P(C|R=r_2, Y=y) = 1 - [1-h(y)]^{r_2}.$$

Now, since $0 \leq h(y) \leq 1$, it is clear that

$$P(C|R=r_1, Y=y) \leq P(C|R=r_2, Y=y).$$

One final note with regard to assumption B6 is in order.

In the case where every cluster has size 1, i.e.,

$$P(R=1) = 1,$$

the probability of sighting a cluster located at a right angle distance y is simply $h(y)$. This is quickly seen by setting $r=1$ in (4.3). Thus, under these circumstances, $h(y)$ has the same interpretation as $g(y)$ defined in Section 2.2.1, that is, $h(y)$ is the conditional probability of sighting an animal at distance y given there is an animal at y .

4.3 General Form of the Likelihood Function

We will use the maximum likelihood procedure to obtain an estimate for D , the animal population density. To obtain the likelihood function, we first need an expression for the probability density function

$$f_{\underline{S}, \underline{Y}, L}(\underline{s}, \underline{y}, \ell)$$

where $\underline{S} = (S_1, S_2, \dots, S_{N_c})$ is the vector of random variables representing the actual number of animals seen in the clusters, $\underline{Y} = (Y_1, Y_2, \dots, Y_{N_c})$ is the vector of random variables representing the right angle distances from the clusters to the transect and L is the random variable representing the total length travelled on the transect to see N_c clusters. Upon writing

$$f_{\underline{S}, \underline{Y}, L}(\underline{s}, \underline{y}, \ell) = f_{\underline{S}|\underline{Y}, L}(\underline{s}|\underline{y}, \ell) f_{\underline{Y}|L}(\underline{y}|\ell) f_L(\ell), \quad (4.4)$$

it is seen that specifying the joint probability density function for $\underline{S}, \underline{Y}$ and L is equivalent to specifying the three functions on the right hand side of (4.4).

The density functions $f_{Y|L}(y|\ell)$ and $f_L(\ell)$ may be derived in a manner analogous to that used in Section 2.2. Let $g_c(y)$ denote the probability of sighting a cluster located at a right angle distance y from the transect, that is

$$g_c(y) = P(\text{observe a cluster} \mid Y=y).$$

Since sighting a cluster located at a distance y is equivalent to observing at least one animal belonging to the cluster, we can write

$$\begin{aligned} g_c(y) &= \sum_{s=1}^{\infty} P(\text{observe } s \text{ animals} \mid Y=y) \\ &= \sum_{s=1}^{\infty} P(S=s \mid Y=y). \end{aligned}$$

Now, for $s \geq 1$,

$$P(S=s \mid Y=y) = \sum_{r=s}^{\infty} P(S=s \mid R=r, Y=y) P(R=r \mid Y=y).$$

By assumption B4, it follows that Y and R are independent random variables. Thus, using (4.1) and (4.2) we get

$$P(S=s \mid Y=y) = \sum_{r=s}^{\infty} \binom{r}{s} [h(y)]^s [1-h(y)]^{r-s} p(r).$$

Therefore,

$$g_c(y) = \sum_{s=1}^{\infty} \sum_{r=s}^{\infty} \binom{r}{s} [h(y)]^s [1-h(y)]^{r-s} p(r). \quad (4.5)$$

Now, according to assumption B6

$$h(0) = 1,$$

so that

$$g_c(0) = \sum_{s=1}^{\infty} p(s) = 1.$$

Therefore, the function $g_c(y)$ plays a role similar to the role of $g(y)$ in Section 2.2. Consequently, by regarding a "cluster" as an "animal," the results of Section 2.2 can be applied to clustered populations in a straightforward manner.

Let $N_c(\ell)$ denote the random variable representing the number of clusters seen when travelling a distance ℓ on the transect. Then, by Theorem 1, $N_c(\ell)$ is a Poisson process with parameter $\theta^*\ell$, where $\theta^*\ell$ is the expected number of clusters seen when travelling along a transect of length ℓ . Also, from Theorem 1 we see that the respective analogs to equations (2.2) and (2.1) are

$$D_c = \frac{\theta^*}{2c^*}, \quad (4.6)$$

where D_c is the density of clusters and

$$c^* = \int_0^\infty g_c(y) dy. \quad (4.7)$$

Furthermore, Theorem 2 gives us the results that L and $\underline{Y}$ are mutually independent random variables, L is distributed as a Gamma random variable with parameters N_c and θ^* and the conditional density of $\underline{Y}$ given $L = \ell$ is

$$f_{\underline{Y}|L}(\underline{y}|\ell) = f_{\underline{Y}}(\underline{y}) = \frac{1}{(c^*)^{N_c}} \prod_{i=1}^{N_c} g_c(y_i). \quad (4.8)$$

Now, assumption B5 implies that the number of animals actually observed in a cluster depends only on the right angle distances to the animals, $\underline{Y}$, and the size of the cluster, R . Thus, $\underline{S}$ is independent of L , and since $\underline{Y}$ is

also independent of L it follows that

$$\begin{aligned} f_{\underline{S}|\underline{Y},L}(\underline{s}|\underline{y},\ell) &= f_{\underline{S}|\underline{Y}}(\underline{s}|\underline{y}) \\ &= \prod_{i=1}^{N_c} P(S_i=s_i | Y_i=y_i). \end{aligned} \quad (4.9)$$

We can now write an expression for the likelihood function $L(\theta^*, p(\cdot), h(\cdot); \underline{s}, \underline{y}, \ell)$. Using (4.4), (4.8) and (4.9) and recalling that L has a Gamma distribution with parameters N_c and θ^* , we obtain

$$\begin{aligned} L(\theta^*, p(\cdot), h(\cdot); \underline{s}, \underline{y}, \ell) &= \prod_{i=1}^{N_c} P(S_i=s_i | Y_i=y_i) \\ &\quad \times \frac{\prod_{i=1}^{N_c} g_c(y_i) (\theta^*)^{N_c} \ell^{N_c-1} e^{-\theta^* \ell}}{(c^*)^{N_c} \Gamma(N_c)}. \end{aligned} \quad (4.10)$$

4.4 Estimation of D when $p(\cdot)$ and $h(\cdot)$ Have Specific Forms

For a clustered population with a cluster density, D_c , the animal population density may be defined as

$$D = D_c v,$$

where $v = E(R)$ is the expected cluster size. Upon using the expression for D_c given in (4.6), we get

$$D = \frac{\theta^* v}{2c^*}, \quad (4.11)$$

so that maximum likelihood estimation of D can be carried out by using (4.11) in the likelihood function presented in (4.10).

Since the random variables $\underline{S}$ and $\underline{Y}$ are independent of L , it is easily seen from (4.10) that the maximum likelihood estimate of θ^* corrected for bias is

$$\theta^* = \frac{N-1}{\underline{c}} . \quad (4.12)$$

However, finding estimates of v and c^* can be quite difficult depending upon the nature of the functions $p(\cdot)$ and $h(\cdot)$. Very likely, one has to resort to some iterative technique such as the Newton-Raphson method (see Korn and Korn, 1968, eqn. (20.2-31)) to solve the likelihood equation.

It is apparent that there exist a wide variety of functions which satisfy the requirements of $p(\cdot)$ and $h(\cdot)$. The appropriate choice in a particular problem depends on the nature of the population under investigation. In this work we will consider the functions

$$p(r) = \frac{\alpha^r e^{-\alpha}}{r!(1-e^{-\alpha})} \quad \alpha > 0, \quad r = 1, 2, \dots, \quad (4.13)$$

and

$$h(y) = e^{-\lambda^* y} \quad \lambda^* > 0, \quad y > 0. \quad (4.14)$$

It is easily seen that $p(\cdot)$ given by (4.13) represents a truncated Poisson distribution. The expected cluster size v is therefore given by

$$v = \frac{\alpha}{1-e^{-\alpha}} . \quad (4.15)$$

The limiting case $\alpha = 0$ corresponds to a population in which the cluster size is 1 with probability 1. Thus, $\alpha = 0$ corresponds to the model in Section 2.2.

The choice for $h(\cdot)$ is based on the fact that when $\alpha = 0$, $h(\cdot)$ may be interpreted as the function $g(\cdot)$ defined in Chapter II. Because

$$g(y) = e^{-\lambda y}$$

seems to be a popular choice for $g(\cdot)$, we feel that

$$h(y) = e^{-\lambda^* y}$$

is a reasonable choice for $h(\cdot)$.

The likelihood function may now be regarded as a function of θ^* , α and λ^* , and maximum likelihood estimation of

$$D = \frac{\theta^* v}{2c^*}$$

may be accomplished by expressing v and c^* as functions of α and λ^* . We have already seen the form of v in equation (4.15). To derive an expression for c^* we proceed as follows. Recall from (4.5) that

$$g_c(y) = \sum_{s=1}^{\infty} P(S=s | Y=y),$$

where

$$P(S=s | Y=y) = \sum_{r=s}^{\infty} \binom{r}{s} [h(y)]^s [1-h(y)]^{r-s} p(r).$$

Now using (4.13) and (4.14) in the above equation, we get

$$\begin{aligned} P(S=s | Y=y) &= \sum_{r=s}^{\infty} \frac{\binom{r}{s} e^{-\lambda^* y s} (1 - e^{-\lambda^* y})^{r-s} \alpha^r e^{-\alpha}}{r! (1 - e^{-\alpha})} \\ &= \frac{(\alpha e^{-\lambda^* y})^s e^{-\alpha}}{s! (1 - e^{-\alpha})} \sum_{r=s}^{\infty} \frac{[\alpha (1 - e^{-\lambda^* y})]^{r-s}}{(r-s)!} \\ &= \frac{(\alpha e^{-\lambda^* y})^s e^{-\alpha e^{-\lambda^* y}}}{s! (1 - e^{-\alpha})}. \end{aligned} \tag{4.16}$$

Then, substituting for $P(S=s|Y=y)$, we get

$$\begin{aligned} g_c(y) &= \frac{e^{-\alpha e^{-\lambda^* y}}}{(1-e^{-\alpha})} \sum_{s=1}^{\infty} \frac{(\alpha e^{-\lambda^* y})^s}{s!} \\ &= \frac{1-e^{-\alpha e^{-\lambda^* y}}}{(1-e^{-\alpha})} . \end{aligned} \quad (4.17)$$

Therefore, using (4.7) and (4.17), we get

$$\begin{aligned} c^* &= \int_0^{\infty} g_c(y) dy \\ &= \int_0^{\infty} \frac{1-e^{-\alpha e^{-\lambda^* y}}}{(1-e^{-\alpha})} dy . \end{aligned} \quad (4.18)$$

To evaluate c^* , note that

$$\int_0^{\infty} (1-e^{-\alpha e^{-\lambda^* y}}) dy = \lim_{\chi \rightarrow \infty} \left\{ \chi - \int_0^{\chi} e^{-\alpha e^{-\lambda^* y}} dy \right\} . \quad (4.19)$$

By letting

$$t = e^{-\lambda^* y}$$

in the integral in the right hand side of (4.19), we can show

$$\begin{aligned} \int_0^{\chi} e^{-\alpha e^{-\lambda^* y}} dy &= \frac{1}{\lambda^*} \int_{e^{-\lambda^* \chi}}^1 \frac{e^{-\alpha t}}{t} dt \\ &= \chi + \frac{1}{\lambda^*} \sum_{j=1}^{\infty} \frac{(-1)^j \alpha^j (1-e^{-\lambda^* \chi})^j}{j \cdot j!} , \end{aligned}$$

and upon substituting into (4.19), we get

$$\int_0^{\infty} (1-e^{-\alpha e^{-\lambda^* y}}) dy = \lim_{\chi \rightarrow \infty} \frac{1}{\lambda^*} \sum_{j=1}^{\infty} \frac{(-1)^j \alpha^j (1-e^{-\lambda^* \chi})^j}{j \cdot j!} .$$

Since the sum above is absolutely convergent, we can take the limit inside the sum and obtain

$$\int_0^{\infty} (1 - e^{-\alpha e^{-\lambda^* y}}) dy = \frac{1}{\lambda^*} \sum_{j=1}^{\infty} \frac{(-1)^j \alpha^j}{j \cdot j!}.$$

Then, using (4.18), it follows that

$$c^* = \frac{a(\alpha)}{\lambda^* (1 - e^{-\alpha})}, \quad (4.20)$$

where

$$a(\alpha) = \sum_{j=1}^{\infty} \frac{(-1)^j \alpha^j}{j \cdot j!}. \quad (4.21)$$

We can now write the likelihood function in terms of θ^* , λ^* and α . Using equations (4.10), (4.16), (4.20) and (4.21), we get

$$\begin{aligned} L(\theta^*, \lambda^*, \alpha; \underline{s}, \underline{y}, \ell) = & \frac{\prod_{i=1}^{N_c} s_i^{-\lambda^*} \prod_{i=1}^{N_c} y_i^{s_i} e^{-\alpha \sum_{i=1}^{N_c} e^{-\lambda^* y_i}}}{(1 - e^{-\alpha})^{N_c} \prod_{i=1}^{N_c} s_i!} \\ & \times \frac{(\lambda^*)^{N_c} \prod_{i=1}^{N_c} (1 - e^{-\alpha e^{-\lambda^* y_i}}) (\theta^*)^{N_c} \ell^{N_c - 1} e^{-\theta^* \ell}}{[a(\alpha)]^{N_c} \Gamma(N_c)}. \end{aligned} \quad (4.22)$$

Using the likelihood function given in (4.22), we can now obtain an estimate for the population density D .

Recall from (4.11) that we can write

$$D = \frac{\theta^* v}{2c^*}.$$

After substituting for c^* and v using equations (4.15), (4.20) and (4.21), the expression for D becomes

$$D = \frac{\theta^* \lambda^* \alpha}{2a(\alpha)} .$$

Thus, an estimate for D would be

$$\hat{D} = \frac{\hat{\theta}^* \hat{\lambda}^* \hat{\alpha}}{2a(\hat{\alpha})} ,$$

where $\hat{\theta}^*$, $\hat{\lambda}^*$ and $\hat{\alpha}$ are maximum likelihood estimates for θ^* , λ^* and α , respectively. As noted earlier in this section, $\underline{S}$ and $\underline{Y}$ are independent of L so that θ^* can be estimated using equation (4.13). However, this still leaves us the problem of estimating λ^* and α . Instead of estimating λ^* and α separately, we can reparameterize the likelihood equation in (4.22) by letting

$$\rho = \frac{\lambda \alpha}{a(\alpha)} ,$$

and

$$\alpha = \alpha .$$

Then, our estimate for D becomes

$$\hat{D}^* = \frac{\hat{\theta}^* \hat{\rho}}{2} . \quad (4.23)$$

The advantage of this reparameterization is that it makes use of the fact that L is independent of both $\underline{S}$ and $\underline{Y}$. Thus, the estimate for D given in (4.23) is now the product of two independent estimates, $\hat{\theta}^*$ which depends on L alone and $\hat{\rho}$ which depends on $\underline{S}$ and $\underline{Y}$. As a result the variance of $\hat{D}^*$ can now be found easily. Using the formula (see Goodman, 1960) for

the variance of the product of two independent estimates, we get

$$\text{Var}(\hat{D}^*) = \frac{1}{4} \left[E^2(\hat{\theta}^*) \text{Var}(\hat{\rho}) + E^2(\hat{\rho}) \text{Var}(\hat{\theta}^*) + \text{Var}(\hat{\theta}^*) \text{Var}(\hat{\rho}) \right] . \quad (4.24)$$

Since L is distributed as a Gamma random variable with parameters N_c and θ^* , exact expressions for $E(\hat{\theta}^*)$ and $\text{Var}(\hat{\theta}^*)$ can be obtained using (2.4) and (2.11), i.e.,

$$E(\hat{\theta}^*) = \theta^*,$$

and

$$\text{Var}(\hat{\theta}^*) = \frac{\theta^{*2}}{N_c - 2} . \quad (4.25)$$

Expressions for the variance and expected value of $\hat{\rho}$ can become quite complicated. An iterative scheme would be needed to find the solutions for ρ and α that would maximize the reparameterized version of the likelihood function given in (4.22). There are computer programs available that can provide maximum likelihood estimates for ρ and α along with numerical approximations for the variance covariance matrix of the estimates. In the next section we will demonstrate the use of one such program with a set of hypothetical data.

4.5 A Worked Example

In this section we will present a worked example to demonstrate the use of a computer program to find the estimate $\hat{D}^*$ and its approximate variance. Because we are not aware of any real data that have been collected according to the sampling plan described in Section 4.2, we shall use an artificial set of data in the example.

Suppose that sampling was continued until $N_c = 25$ clusters were sighted, and that a transect length of $\ell = 25$ miles was needed to sight the 25 clusters. Suppose further that the observed right angle distances and the cluster sizes were as follows, where the first number in the pair is the right angle distance, y , measured in yards and the second number in the pair is the corresponding cluster size, s :

(1,1), (3,2), (7,1), (10,1), (2,3)
 (5,5), (4,1), (7,2), (15,1), (22,1)
 (6,1), (3,6), (2,1), (12,1), (28,3)
 (9,2), (18,1), (36,7), (17,6), (5,1)
 (4,1), (3,1), (8,2), (3,4), (13,1).

As noted in Section 4.4, an estimate for θ^* is

$$\hat{\theta}^* = \frac{N_c - 1}{\ell} = .96,$$

and an estimate for the variance of $\hat{\theta}^*$ is

$$\hat{\text{Var}}(\hat{\theta}^*) = \frac{\hat{\theta}^{*2}}{N_c - 2} = .0401.$$

In order to estimate ρ , the reparameterized version of the likelihood function given in (4.22) will have to be maximized. The Fortran subroutine ZXMIN, found in IMSL (1979) may be used for this purpose. This program uses a quasi-Newton iterative procedure to find the minimum of a function. Thus, we first need to take the negative of the likelihood equation before we can use this subroutine to our advantage.

On output, this subroutine not only provides the values at which the function is minimized, but also provides numerical estimates for the second partial derivatives of the function evaluated at the minimization point. Thus, when used with the negative of the likelihood function this program will provide the maximum likelihood estimates, $\hat{\rho}$ and $\hat{\alpha}$, as well as the matrix of negative second partial derivatives of the likelihood, $L(\cdot)$, evaluated at $\hat{\rho}$ and $\hat{\alpha}$. We will denote this matrix by

$$V^{-1} = - \begin{pmatrix} \frac{\partial^2 \ln L(\cdot)}{\partial \alpha^2} & \frac{\partial^2 \ln L(\cdot)}{\partial \alpha \partial \rho} \\ \frac{\partial^2 \ln L(\cdot)}{\partial \alpha \partial \rho} & \frac{\partial^2 \ln L(\cdot)}{\partial \rho^2} \end{pmatrix}_{\substack{\alpha = \hat{\alpha} \\ \rho = \hat{\rho}}}.$$

For our data, the use of the subroutine ZXMIN with initial values $\alpha_I = 2.24$ and $\rho_I = .16$ yielded

$$\hat{\alpha} = 2.844,$$

$$\hat{\rho} = .0907,$$

and

$$V^{-1} = \begin{pmatrix} 7.687 & -161.229 \\ -161.229 & 5098.985 \end{pmatrix} .$$

The initial value used for α was the mean of the observed cluster sizes, i.e.,

$$\alpha_I = \frac{\sum_{i=1}^{25} s_i}{25} = \bar{s} .$$

Since our model does not assume all animals belonging to a cluster are seen, $\bar{s}$ would underestimate the expected cluster size, i.e.,

$$\bar{s} < E(R) = \frac{\alpha}{1-e^{-\alpha}} .$$

Thus, $\bar{s}$ seems to be a good starting value for α .

In choosing an initial value for ρ , first recall that

$$\rho = \frac{\lambda^* \alpha}{a(\alpha)} ,$$

where $a(\alpha)$ is given in equation (4.21). Since our initial value for α is $\bar{s}$, all we need is a starting value for λ^* . If every animal in the cluster was seen with probability 1, the density of clusters would be estimated by the method described in Chapter II. In this case, the maximum likelihood estimate for λ^* would be $1/\bar{y}$ where

$$\bar{y} = \frac{\sum_{i=1}^{25} y_i}{25} .$$

Thus, as the initial value for ρ we used

$$\rho_I = \frac{\bar{s}}{\bar{y}a(\bar{s})}.$$

The estimate for the density can now be calculated. Using (4.23) and substituting the values we obtained for $\hat{\theta}$ and $\hat{\rho}$, we get

$$\hat{D}^* = 76.7 \text{ animals/square mile.}$$

Now if we can obtain a large sample approximation for the variance of $\hat{\rho}$, then we can use (4.24) as an approximation for the variance of $\hat{D}^*$. Now, under the usual regularity conditions, V^{-1} will be a large sample approximation to the inverse of the variance-covariance matrix of $\hat{\alpha}$ and $\hat{\rho}$. Furthermore, the approximate variance of $\hat{D}^*$ can be obtained from equation (4.24) after substituting the element in the matrix corresponding to the approximate variance of $\hat{\rho}$ along with the other appropriate quantities. Straightforward calculations show that

$$\sqrt{\text{Var}(\hat{D}^*)} \doteq 26.2 \text{ animals/square mile.}$$

The use of this Fortran subroutine required a minimal amount of programming to enter the appropriate likelihood function. It was run using the computer facilities of the Northeast Regional Data Center located in Gainesville, Florida. Less than two seconds of CPU time was needed for the estimates to converge to values that agreed to four significant digits on two successive iterations.

BIBLIOGRAPHY

- Anderson, D. R., Laake, J. L., Crain, B. R., and Burnham, K. P. (1976), Guidelines for Line Transect Sampling of Biological Populations, Logan: Utah Cooperative Wildlife Research Unit.
- Anderson, D. R., and Pospahala, R. S. (1970), "Correction of Bias in Belt Transect Studies of Immobile Objects," Journal of Wildlife Management, 34, 141-146.
- Barr, A. J., Goodnight, J. H., Sall, J. P., and Helwig, J. T. (1976), A User's Guide to SAS 76, Raleigh: SAS Institute.
- Bhat, U. N. (1972), Elements of Applied Stochastic Processes, New York: John Wiley & Sons.
- Burnham, K. P., and Anderson, D. R. (1976), "Mathematical Models for Nonparametric Inferences from Line Transect Data," Biometrics, 32, 325-336.
- Crain, B. R., Burnham, K. P., Anderson, D. R., and Laake, J. L. (1978), A Fourier Series Estimator of Population Density for Line Transect Sampling, Logan: Utah State University Press.
- Eberhardt, L. L. (1968), "A Preliminary Appraisal of Line Transects," Journal of Wildlife Management, 32, 82-88.
- Gates, C. E., Marshall, W. H., and Olson, D. P. (1968), "Line Transect Method of Estimating Grouse Population Densities," Biometrics, 24, 135-145.
- Goodman, L. A. (1960), "On the Exact Variance of Products," Journal of the American Statistical Association, 55, 708-713.
- Hayne, D. W. (1949), "An Examination of the Strip Census Method for Estimating Animal Populations," Journal of Wildlife Management, 13, 145-157.
- IMSL (1979), The IMSL Library, Seventh ed., Vol. 3, Houston: International Mathematical and Statistical Libraries, Inc.
- Korn, G. A., and Korn, T. M. (1968), Mathematical Handbook for Scientists and Engineers, Second ed., New York: McGraw-Hill.

- Leopold, A. (1933), Game Management, New York: Charles Scribner's Sons.
- Lindgren, B. W. (1968), Statistical Theory, Second ed., New York: Macmillan.
- Loftsgaarden, D. O., and Quesenberry, C. P. (1965), "A Non-parametric Estimate of a Multivariate Density Function," Annals of Mathematical Statistics, 36, 1049-1051.
- Pielou, E. C. (1969), An Introduction to Mathematical Ecology, New York: John Wiley & Sons.
- Pollock, K. H. (1978), "A Family of Density Estimators for Line Transect Sampling," Biometrics, 34, 475-478.
- Robinette, W. L., Jones, D. A. Gashwiler, J. S., and Aldous, C. M. (1954), "Methods for Censusing Winter-Lost Deer," North American Wildlife Conference Transactions, 19, 511-524.
- Robinette, W. L., Loveless, C. M., and Jones, D. A. (1974), "Field Tests of Strip Census Methods," Journal of Wildlife Management, 38, 81-96.
- Seber, G. A. F. (1973), The Estimation of Animal Abundance and Related Parameters, London: Griffin.
- Sen, A. R., Tournigny, J., and Smith, G. E. J. (1974), "On the Line Transect Sampling Method," Biometrics, 30, 329-340.
- Smith, M. H., Gardner, R. H., Gentry, J. B., Kaufman, D. W., and O'Farrel, M. H. (1975), Small Mammals: Their Productivity and Population Dynamics, International Biological Program.
- Webb, W. L. (1942), "Notes on a Method of Censusing Snowshoe Hare Populations," Journal of Wildlife Management, 6, 67-69.

BIOGRAPHICAL SKETCH

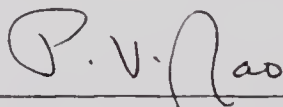
John Anthony Ondrasik was born on August 17, 1951, in New Brunswick, New Jersey. Shortly thereafter his parents moved to Palmerton, Pennsylvania, where he grew up and attended high school. After graduation in June, 1969, he entered Bucknell University in Lewisburg, Pennsylvania, and received the degree of Bachelor of Science with a major in mathematics in June, 1973.

It was during his studies at Bucknell that he became interested in statistics through the influence of the late Professor Paul Benson. In September, 1973, he matriculated in the graduate school at the University of Florida and received the degree Master of Statistics in 1975.

While pursuing his graduate studies, he worked for the Department of Statistics as an assistant in their biostatistics consulting unit. In November, 1978, he accepted the position of biostatistician with Boehringer Ingelheim, Ltd.

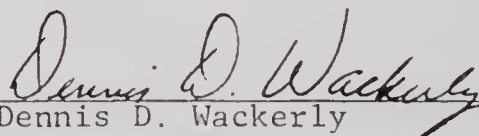
John Ondrasik is married to the former Anntoinette M. Lucia. Currently they reside in Danbury, Connecticut.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



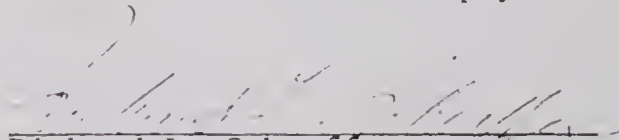
Peiaver V. Rao, Chairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



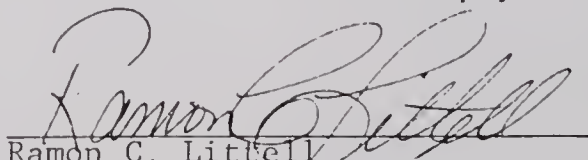
Dennis D. Wackerly
Associate Professor
of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Richard L. Scheaffer
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Ramon C. Littell
Associate Professor
of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Wayne R. Marion
Assistant Professor
of Forest Resources and
Conservation

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December 1979

Dean Graduate School